

When do cross-sectional asset pricing factors span the SDF?

Serhiy Kozak¹ Stefan Nagel²

¹University of Maryland

²University of Chicago, NBER, CEPR, and CESifo

Characteristics-based factor models

- ▶ Stock characteristics

$$\mathbf{X}_t = \begin{pmatrix} 1 & sz_{1,t} & bm_{1,t} & mom_{1,t} & \dots \\ 1 & sz_{2,t} & \dots & \dots & \\ \dots & & & & \\ 1 & sz_{N,t} & \dots & \dots & \end{pmatrix}$$

$(N \times J)$

- ▶ Excess returns \mathbf{z}_{t+1}
 $(N \times 1)$

- ▶ Cross-sectional stock return predictability:

$$\mathbb{E}[\mathbf{z}_{t+1} | \mathbf{X}_t] = \mathbf{X}_t \boldsymbol{\phi}$$

- ▶ Characteristics-based factor models: SDF with factors

$$\mathbf{f}_{t+1} = f(\mathbf{z}_{t+1}, \mathbf{X}_t), \quad K \leq J$$

$(K \times 1)$

Heuristic equity factor construction methods

- ▶ Common factor construction methods: N stocks to $J < N$ factors
 - ▶ Sorted factors: $\mathbf{X}'_t \mathbf{z}_{t+1}$ with characteristics transformed into bin indicators (e.g., FF 1993)
 - ▶ Univariate factors: $\mathbf{X}'_t \mathbf{z}_{t+1}$ (e.g., KNS 2020)
 - ▶ OLS factors: $(\mathbf{X}'_t \mathbf{X}_t)^{-1} \mathbf{X}'_t \mathbf{z}_{t+1}$ (e.g., FF 2020; BARRA)
- ▶ Objective of such reduced-form factor model construction: span MV frontier
- ▶ Generally, MV efficient portfolio depends on covariance matrix, but these methods don't use it
- ▶ What are the necessary and sufficient conditions for these approaches to yield factors that span the MV frontier?

Heuristic factor hedging

- ▶ Literature recognized that heuristic factor construction methods are imperfect: factors contaminated with unpriced risk
- ▶ Heuristic attempts to hedge unpriced risks and thereby construct cleaner factors closer to spanning MV frontier without using full covariance matrix
- ▶ Example: Daniel, Mota, Rottke, Santos (DMRS) (2020)
- ▶ Approach seems to have some success empirically: yields factors with higher Sharpe ratios
- ▶ What are sufficient conditions for this approach to recover factors that span the MV frontier?
 - ▶ Empirical SR improvements due to hedging quantify the efficiency loss of heuristic factors.

Heuristic dimension reduction

- ▶ Dimension reduction methods: Compress information in J characteristics-based factors into $K < J$ factors (without having to invert a $J \times J$ covariance matrix)
- ▶ Heuristic approaches
 - ▶ PCA on characteristics-based factors (KNS 2020)
 - ▶ Instrumented PCA (Kelly, Pruitt, and Su 2019)
 - ▶ Projected PCA (Kim, Korajczyk, and Neuhierl 2019)
- ▶ What are the necessary and sufficient conditions for dimension reduction to be possible and for these methods to yield factors that span the MV frontier?

Outline

1. Conditions for characteristics-based factors to span the MV frontier
 - ▶ Heuristic factors
 - ▶ Factor hedging
 - ▶ Iterated factor hedging
2. Dimension reduction
3. Empirics

Setup

- ▶ \mathbf{X}_t includes a column of ones and is observable to econometrician
- ▶ Conditional moments

$$\boldsymbol{\Sigma}_t = \text{var}(\mathbf{z}_{t+1} | \mathbf{X}_t), \quad \boldsymbol{\mu}_t = \mathbb{E}[\mathbf{z}_{t+1} | \mathbf{X}_t],$$

- ▶ SDF in the span of individual stock excess returns:

$$M_{t+1} = 1 - \mathbf{b}'_t (\mathbf{z}_{t+1} - \boldsymbol{\mu}_t), \quad \mathbf{b}_t = \boldsymbol{\Sigma}_t^{-1} \boldsymbol{\mu}_t,$$

for which

$$\mathbb{E}[M_{t+1} \mathbf{z}_{t+1} | \mathbf{X}_t] = 0$$

Factors

- ▶ Factors are generally constructed with an $N \times J$ portfolio weight matrix \mathbf{W}_t :

$$\mathbf{f}_{t+1} = \mathbf{W}'_t \mathbf{z}_{t+1},$$

with $\boldsymbol{\mu}_{f,t} = \mathbf{W}'_t \boldsymbol{\mu}_t$ and $\boldsymbol{\Sigma}_{f,t} = \mathbf{W}'_t \boldsymbol{\Sigma}_t \mathbf{W}_t$.

- ▶ Under which conditions do different specifications of \mathbf{W}_t produce factors that span the **conditional** MV frontier, which means SDF can be written as

$$M_{t+1} = 1 - \boldsymbol{\mu}'_{f,t} \boldsymbol{\Sigma}_{f,t}^{-1} (\mathbf{f}_{t+1} - \boldsymbol{\mu}_{f,t}) \quad ?$$

When do factors span the conditional MV frontier?

Lemma 1

The maximum squared conditional Sharpe ratio of the factors $\mathbf{f}_{t+1} = \mathbf{W}'_t \mathbf{z}_{t+1}$ is equal to the maximum squared conditional Sharpe Ratio of the individual assets, i.e.,

$$\boldsymbol{\mu}'_t \boldsymbol{\Sigma}_t^{-1} \boldsymbol{\mu}_t = \boldsymbol{\mu}'_t \mathbf{W}_t (\mathbf{W}'_t \boldsymbol{\Sigma}_t \mathbf{W}_t)^{-1} \mathbf{W}'_t \boldsymbol{\mu}_t \quad (1)$$

if and only if

$$\boldsymbol{\mu}_t = \boldsymbol{\Sigma}_t \mathbf{W}_t \mathbf{b}_t \quad (2)$$

for some $J \times 1$ vector \mathbf{b}_t .

Intuition: ER must be proportional to cov of returns and factors, $\boldsymbol{\Sigma}_t \mathbf{W}_t$. Equivalently, SDF risk prices $\boldsymbol{\Sigma}_t^{-1} \boldsymbol{\mu}_t$ must be spanned by factor weights \mathbf{W}_t .

Linearity of expected returns in characteristics

Assumption 1

For some $J \times 1$ vector ϕ ,

$$\mu_t = \mathbf{X}_t \phi$$

- ▶ When do the researcher's characteristics-based factors fail to span the SDF even though conditional expected returns are perfectly linear in \mathbf{X}_t ?
- ▶ Conceptually not restrictive as \mathbf{X}_t can include nonlinear transformations/expansions of characteristics and state variables
- ▶ In empirical applications, assumption becomes substantive once specific \mathbf{X}_t chosen

MV efficient factors: GLS factors

Proposition 1

Assumption 1 is equivalent to the statement that characteristics-based factors

$$\mathbf{f}_{t+1} = \mathbf{S}'_t \mathbf{X}'_t \boldsymbol{\Sigma}_t^{-1} \mathbf{z}_{t+1},$$

are conditionally MV efficient.

- ▶ Example: $\mathbf{S}_t = (\mathbf{X}'_t \boldsymbol{\Sigma}_t^{-1} \mathbf{X}_t)^{-1}$. Then factors are **GLS cross-sectional regression slopes**

$$\mathbf{f}_{t+1} = (\mathbf{X}'_t \boldsymbol{\Sigma}_t^{-1} \mathbf{X}_t)^{-1} \mathbf{X}'_t \boldsymbol{\Sigma}_t^{-1} \mathbf{z}_{t+1}$$

and **factor betas = characteristics**

$$\boldsymbol{\beta}_t = \mathbf{X}_t$$

Horseraces of characteristics vs. covariances

- ▶ With GLS factors prescribed by theory, no difference between direct linear prediction of returns by \mathbf{X}_t and factor pricing model with factors constructed based on \mathbf{X}_t
- ▶ Horseraces of characteristics vs. covariances w.r.t. ad-hoc factors [e.g., in Daniel and Titman (1997) and Davis, Fama, and French (2000)] may produce a pricing wedge just because heuristic factors \neq GLS factors
- ▶ Such horseraces therefore cannot discriminate between “risk-based” vs. “behavioral” explanations

Example: One characteristic

- ▶ Suppose \mathbf{x}_t is the only priced characteristic, $\boldsymbol{\mu}_t = \mathbf{x}_t\phi$. Let $\mathbf{x}_t'\mathbf{x}_t = 1$.
- ▶ We can always write the conditional covariance matrix as

$$\boldsymbol{\Sigma}_t = \mathbf{x}_t\psi_{1,t}\mathbf{x}_t' + \mathbf{U}_t\boldsymbol{\Omega}_t\mathbf{U}_t' \quad (3)$$

- ▶ Consider a heuristic factor constructed as $\mathbf{f}_{t+1} = \mathbf{x}_t'\mathbf{z}_{t+1}$. The covariances of this factor with individual stocks are $\boldsymbol{\Sigma}_{zf,t} = \boldsymbol{\Sigma}_t\mathbf{x}_t$, and hence

$$\boldsymbol{\Sigma}_{zf,t} = \mathbf{x}_t\psi_{1,t} + \underbrace{\mathbf{U}_t\boldsymbol{\Omega}_t\mathbf{U}_t'\mathbf{x}_t}_{\text{Unpriced risk contamination}} \quad (4)$$

- ▶ Covariances are not linear in \mathbf{x}_t due to the second term, unless $\mathbf{U}_t'\mathbf{x}_t = \mathbf{0}$.
- ▶ For \mathbf{f}_{t+1} to correctly price individual stocks, \mathbf{x}_t must be orthogonal to loadings on systematic factors other than \mathbf{f}_{t+1} that appear in the covariance matrix.

Heuristic factors: OLS factors and rotations thereof

Proposition 2

Suppose Assumption 1 holds. Then factors

$$\mathbf{f}_{t+1} = \mathbf{S}'_t \mathbf{X}'_t \mathbf{z}_{t+1},$$

are conditionally MV efficient if and only if there exist conformable matrices Ψ_t , Ω_t , and a matrix \mathbf{U}_t for which

$$\mathbf{U}'_t \mathbf{X}_t = \mathbf{0},$$

such that

$$\Sigma_t = \mathbf{X}_t \Psi_t \mathbf{X}'_t + \mathbf{U}_t \Omega_t \mathbf{U}'_t.$$

- ▶ $\mathbf{U}'_t \mathbf{X}_t = \mathbf{0}$ prevents contamination of \mathbf{f}_{t+1} with unpriced risk
- ▶ Examples:
 - ▶ OLS factors (FF 2018): $\mathbf{S}_t = (\mathbf{X}'_t \mathbf{X}_t)^{-1}$
 - ▶ Univariate factors (KNS 2020): $\mathbf{S}_t = \mathbf{I}$

Example: Two correlated characteristics

► Suppose $\mathbf{X}_t = (\mathbf{x}_t \quad \mathbf{y}_t)$ with $\mathbf{\Psi}_t$ diagonal and $\mathbf{U}'_t \mathbf{X}_t = 0$.

► Then

$$\mathbf{\Sigma}_t = \mathbf{x}_t \psi_{1,t} \mathbf{x}'_t + \mathbf{y}_t \psi_{2,t} \mathbf{y}'_t + \mathbf{U}_t \mathbf{\Omega}_t \mathbf{U}'_t. \quad (5)$$

► Even if only \mathbf{x}_t relevant for expected returns, i.e., $\boldsymbol{\mu}_t = \mathbf{x}_t \phi$, the covariances of individual stocks with this factor are contaminated by unpriced risk:

$$\Sigma_{zf,t} = \mathbf{x}_t \psi_{1,t} + \underbrace{\mathbf{y}_t \psi_{2,t} \mathbf{y}'_t \mathbf{x}_t}_{\text{Unpriced risk contamination}}. \quad (6)$$

► Inclusion of \mathbf{y}_t in \mathbf{X}_t purges unpriced risk and makes the OLS factors mean-variance efficient, even though \mathbf{x}_t alone fully explains expected returns.

Benefits of broadening set of characteristics

- ▶ Including more characteristics in factor construction makes condition

$$\Sigma_t = \mathbf{X}_t \Psi_t \mathbf{X}_t' + \mathbf{U}_t \Omega_t \mathbf{U}_t', \quad \mathbf{U}_t' \mathbf{X}_t = \mathbf{0}$$

more likely to hold, at least approximately: If \mathbf{X}_t spans major sources of covariances, $\mathbf{U}_t' \mathbf{X}_t$ quantitatively small, even if nonzero.

- ▶ Even if a characteristic does not contribute to expected return variation, as long as it helps span major sources of covariances, it helps OLS factors to span the SDF.
- ▶ Literature has focused almost exclusively only on characteristics that contribute to expected return variation

Hedging unpriced risks

- ▶ Ideal, but infeasible in practice: Invert Σ_t to construct GLS factors to completely avoid unpriced risk contamination
- ▶ Hedging methods use **some** information from covariance matrix to try to get heuristic factors closer to GLS factors without inverting full Σ_t .
- ▶ Heuristic hedging methods proposed in literature: Not clear whether and why they work
- ▶ Our results related to older literature in econometrics on **partially GLS** estimators

Example: Hedging with two correlated characteristics

- ▶ Continuing the previous example, consider the hedged factors $\mathbf{f}_{t+1} = \mathbf{h}'_t \mathbf{z}_{t+1}$ where

$$\mathbf{h}_t = \mathbf{x}_t - \mathbf{y}_t(\mathbf{y}'_t \mathbf{y}_t)^{-1} \mathbf{y}'_t \mathbf{x}_t. \quad (7)$$

- ▶ Removing the projection of \mathbf{x}_t on \mathbf{y}_t from \mathbf{x}_t in the hedged factor portfolio weights removes this unpriced risk: $\mathbf{y}'_t \mathbf{h}_t = 0$.
- ▶ Covariances of individual stocks with the hedged factor are then proportional to \mathbf{x}_t and hence perfectly aligned with $\boldsymbol{\mu}_t$.
- ▶ ...but \mathbf{y}_t is not observable...

Example: DMRS approach with two characteristics

1. Start with $f_{t+1} = \mathbf{x}'_t \mathbf{z}_{t+1}$ and recall that

$$\Sigma_{zf,t} = \mathbf{x}_t \times \text{scalar} + \mathbf{y}_t \times \text{scalar}$$

2. $\Sigma_{zf,t}$ residualized w.r.t $\mathbf{x}_t = \mathbf{v}_t \times \text{scalar}$
 - ▶ where $\mathbf{v}_t = \mathbf{y}_t - \mathbf{x}_t(\mathbf{x}'_t \mathbf{x}_t)^{-1} \mathbf{x}'_t \mathbf{y}_t$

3. Hedging factor weights = $\mathbf{v}_t \times \text{scalar}$
4. Stocks' covariance with hedging factors = $\mathbf{y}_t \times \text{scalar}$
5. Residualizing \mathbf{x}_t w.r.t. these covariances yields

$$\mathbf{h}_t = \mathbf{x}_t - \mathbf{y}_t(\mathbf{y}'_t \mathbf{y}_t)^{-1} \mathbf{y}'_t \mathbf{x}_t$$

6. Stocks' covariances with $f_{h,t+1} = \mathbf{h}'_t \mathbf{z}_{t+1}$ are $\mathbf{x}_t \times \text{scalar}$ ✓

Iterated hedging

- ▶ No reason to stop after one round of hedging: can repeat the procedure with hedged factors from first round now the starting point for the second round
- ▶ Each additional hedging round
 - ▶ uses additional information from the covariance matrix
 - ▶ allows for J additional components in Σ_t correlated with \mathbf{X}_t
- ▶ Procedure should converge to GLS factors
- ▶ In practice, with estimated moments, decay in Sharpe ratios may set in as hedged factors more and more contaminated by estimation noise

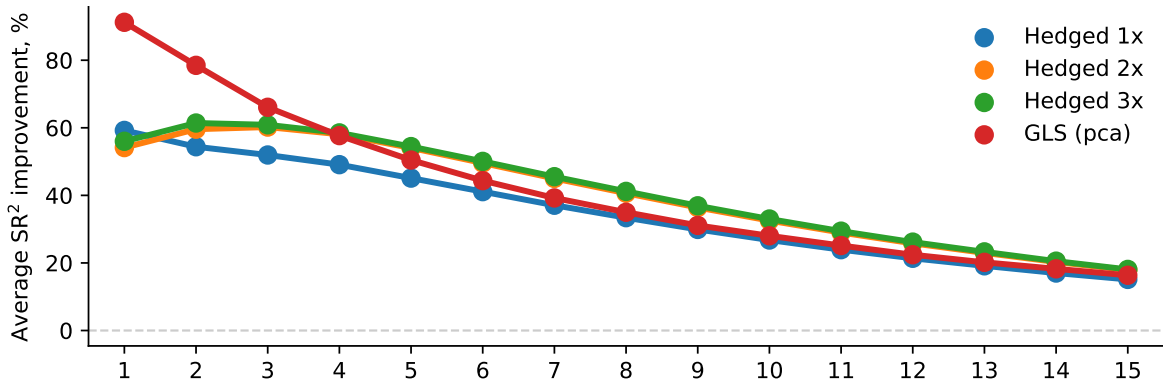
Dimensionality reduction

- ▶ Conditions on Σ_t , such that one can summarize pricing info in J characteristics-based factors in a smaller number of $K < J$ factors (without having to invert full Σ_t)?
- ▶ We provide necessary, not just sufficient conditions for dimension-reduction to be possible
- ▶ Under these conditions: equivalence of IPCA (Kelly, Pruitt, and Su 2019) and PPCA (Kim, Korajczyk, and Neuhierl 2019) to simple PCA on certain factor portfolios

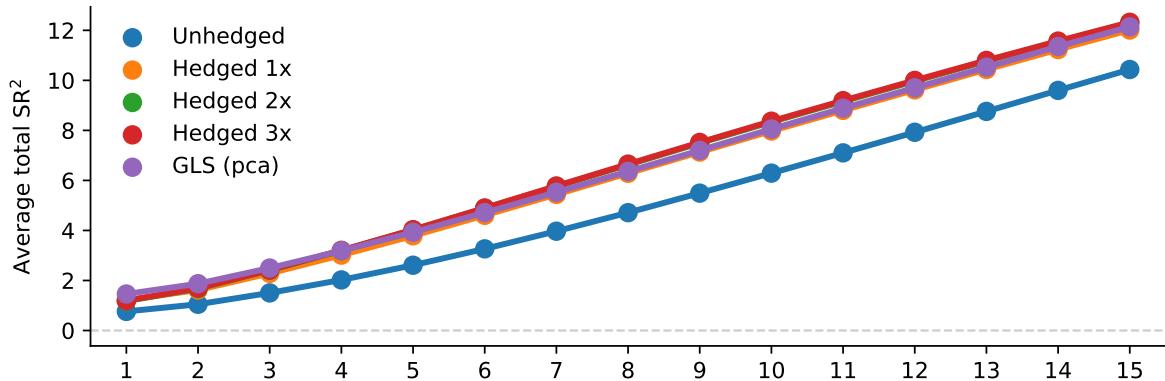
Empirical analysis

- ▶ How close do OLS factors get to conditional MV efficiency?
 - ▶ Compare with approximate GLS factors
 - ▶ Alternative: Examine gain from hedging unpriced risks
- ▶ Are gains from hedging/approximate GLS smaller when many characteristics are used, as our theoretical results suggest?
- ▶ Data
 - ▶ 34 characteristics
 - ▶ Microcaps excluded
 - ▶ 1972-2021 (Est. 1972-2005; OOS 2005-2021)

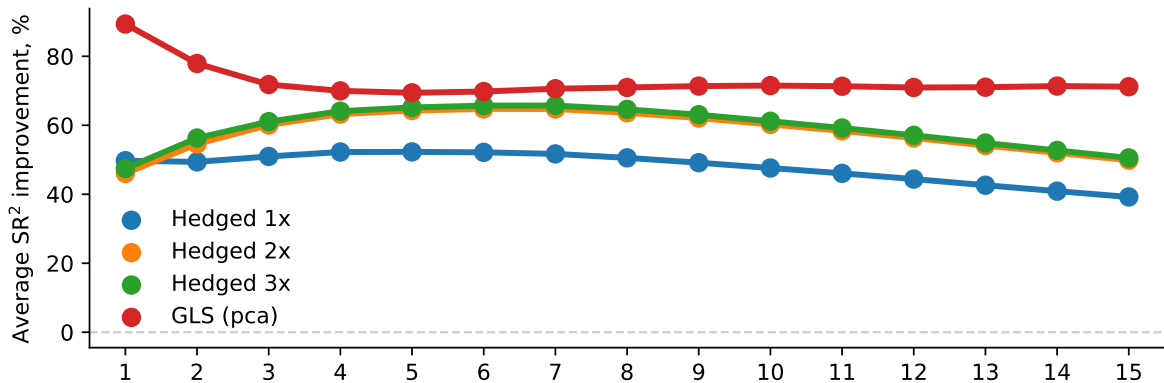
IS improvement in squared SR due to iterative hedging: OLS factors



IS squared SR with and without hedging: OLS factors



IS improvement in squared SR due to iterative hedging: Univariate factors



Maximum squared in-sample Sharpe ratios of one-factor OLS models

	OLS	Hedged n times			GLS	
		1	2	3	PCA	Char.
Size	0.4	0.7*	0.6	0.6	0.8**	1.4**
Gross Profitability	0.5	1.0**	0.9*	0.9*	1.1**	1.4**
F-score	0.8	1.6**	1.9**	1.9**	2.2**	1.5*
Share Repurchases	0.7	1.3**	1.4*	1.5**	1.4**	1.3*
Net Issuance (A)	1.1	1.9**	2.0**	2.2**	1.9**	2.0**
Asset Growth	0.9	1.2*	1.3*	1.4*	1.5**	3.2**
Return on Assets (A)	0.4	0.8*	0.7	0.7*	1.1**	1.2**
Industry Momentum	1.0	1.6*	1.6*	1.5*	2.0**	3.5**
Momentum (12m)	0.7	1.1*	1.1*	1.1*	1.6**	3.6**
Value (M)	0.6	0.9*	0.9*	0.9*	1.1**	1.6**
Net Issuance (M)	1.1	1.9**	1.9*	1.9*	2.4**	2.3**
Short-Term Reversals	0.7	1.6**	1.6**	1.6**	2.1**	5.4**
Industry Rel. Reversals	1.5	2.6**	2.6**	2.7**	3.4**	8.1**
ER	7.1	9.1*	9.6**	9.7**	11.3**	13.5**
Average	0.9	1.4	1.3	1.4	1.6	2.3

Maximum squared Sharpe ratios of hedged factors

	Unhedged	Hedged n times				
		1	2	3	4	5
		In-sample				
Univariate	13.8	16.4*	17.2**	17.3**	17.4**	17.4**
Orthonormal	18.0	20.1*	20.3*	20.4*	20.5*	20.5*
OLS	21.3	21.8	21.5	21.6	21.6	21.6
		Out-of-sample				
Univariate	1.3	1.5	2.1	2.3	2.4	2.5
Orthonormal	3.4	3.1	3.4	3.6	3.6	3.7
OLS	4.0	3.8	3.9	4.1	4.1	4.2

Dimensionality reduction: benchmarking portfolio sorts

	1	2	3	4	5	6	7	8	9	10	11	12
	In-sample											
SCS	0.2	0.6	0.9	1.2	3.1	3.1	3.1	4.4	4.7	4.7	7.9	8.1
IPCA	0.3	1.3	4.1	4.5	7.0	7.7	11.9	12.6	13.7	14.4	14.8	15.3
PPCA	0.3	0.3	0.7	2.5	8.3	8.3	8.7	12.0	12.0	13.2	13.2	13.3
IPCA (GLS)	0.6	1.3	11.1	10.9	12.0	12.9	16.4	16.8	16.7	16.5	16.3	16.4
	Out-of-sample											
SCS	0.1	0.2	0.4	0.5	0.4	0.3	0.3	0.6	0.8	0.8	1.6	1.5
IPCA	0.3	0.1	0.7	0.8	1.0	1.1	2.1	2.2	2.6	3.1	3.5	3.8
PPCA	0.2	0.2	0.4	1.0	1.6	1.3	1.2	3.0	2.4	3.2	3.1	3.1
IPCA (GLS)	0.4	0.2	2.7	2.2	2.8	2.8	4.7	4.8	3.8	3.8	3.8	3.7

Dimensionality reduction: hedging latent factors (part I)

	1	2	3	4	5	6	7	8	9	10	11	12
SCS												
Unhedged	0.2	0.6	0.9	1.2	3.1	3.1	3.1	4.4	4.7	4.7	7.9	8.1
Hedged 1x	0.1	0.6	1.3	1.7	7.4**	7.5**	7.6**	9.1**	9.2**	9.3**	10.9**	10.9**
Hedged 2x	0.1	0.6	1.2	1.8*	7.6**	7.8**	7.9**	9.6**	9.7**	9.8**	11.4**	11.5**
Hedged 3x	0.1	0.6	1.3*	1.8*	7.6**	7.8**	7.8**	9.6**	9.7**	9.8**	11.5**	11.6**
IPCA												
Unhedged	0.3	1.3	4.1	4.5	7.0	7.7	11.9	12.6	13.7	14.4	14.8	15.3
Hedged 1x	0.4	1.4	4.6	5.1	6.8	7.7	12.7	13.4	15.1	15.8	16.0	16.2
Hedged 2x	0.3	1.4	5.0*	5.6*	7.4	8.2	13.3	13.7	15.1	15.9	16.0	16.3
Hedged 3x	0.4	1.5	4.9	5.6*	7.4	8.2	13.4	13.7	15.1	15.8	15.9	16.2

Dimensionality reduction: hedging latent factors (part II)

	1	2	3	4	5	6	7	8	9	10	11	12
	PPCA											
Unhedged	0.3	0.3	0.7	2.5	8.3	8.3	8.7	12.0	12.0	13.2	13.2	13.3
Hedged 1x	0.3	0.3	1.1*	3.9**	12.4**	12.7**	13.0**	14.5**	15.0**	15.8**	16.0**	16.0**
Hedged 2x	0.3	0.3	1.1	4.1**	12.6**	12.9**	13.3**	14.8**	15.0**	15.7*	16.0**	16.1**
Hedged 3x	0.4	0.3	1.1*	4.1**	12.6**	12.8**	13.2**	14.8**	15.0**	15.6*	16.0**	16.0**
	IPCA (GLS)											
Unhedged	0.6	1.3	11.1	10.9	12.0	12.9	16.4	16.8	16.7	16.5	16.3	16.4
Hedged 1x	0.6	1.3	11.1	10.8	12.0	12.8	16.5	16.9	16.8	16.5	16.3	16.3
Hedged 2x	0.6	1.3	11.0	10.7	11.9	12.8	16.5	17.0	16.9	16.6	16.4	16.4
Hedged 3x	0.6	1.3	11.0	10.7	11.9	12.8	16.5	16.9	16.8	16.5	16.3	16.4

Conclusion

- ▶ Heuristic factor models avoid need for knowing Σ_t , but span MV frontier under certain conditions on Σ_t
 - ▶ Conditions more likely to hold with inclusion of more characteristics. Characteristics can help even if related only to Σ_t but not expected returns
 - ▶ (Iterated) hedging of unpriced risks allows weakening of conditions. Uses partial information about Σ_t .
 - ▶ Hedging and GLS factor constructions improve MV efficiency of small-scale factor models empirically, both IS and OOS
 - ▶ More so for Univariate factors, suggesting they are less efficient
- ▶ Dimensionality reduction possible under joint conditions on Σ_t and expected returns
 - ▶ IPCA and PPCA closely linked to PCA on simple characteristics factors. All equivalent if characteristics orthonormalized
 - ▶ Empirically, latent factor models perform quite differently depending on how their factors are constructed; might benefit from hedging