Factor Models, Machine Learning, and Asset Pricing

Stefano Giglio, Bryan Kelly, and Dacheng Xiu

Presented by: Zhen Long

Introduction

- Factor models offer a parsimonious statistical description of returns' cross-sectional dependence structure.
- The arbitrage pricing theory(APT) ties factors to fundamental economic concepts, such as risk exposure and risk premia.
- → unobservable asset risk premia are difficult to pinpoint:
  - Return variation dominated by unforecastable news
  - T<<N
  - Function ambiguity
- Variable selection and dimension reduction paradigms
  - Sort portfolio: cope with nonlinearity, low signal-noise ratio, curse of dimensionality
  - →data-driven solutions

Contents

Model specification - static factor models

- Simplest form: $$r_t = \mathrm{E}(r_t) + \beta v_t + u_t,$$

  β is an $N \times K$ matrix of factor exposures, v_t is a $K \times 1$ vector of factor innovations.

  $$\mathrm{E}(r_t) = \alpha + \beta\gamma,$$

  γ is a $K \times 1$ vector of risk premia and α is an $N \times 1$ vector of pricing errors.

- Three framework:
  - Factors are known and observable (eg. Market index)
  - All factors and exposures are latent (eg. Political risk)
  - Exposures are observable but the factors are latent (eg. PE ratio)

- Risk exposure changes over time; derivatives and bonds→

Model specification- conditional factor models

- The conditional factor model can be specified as

$$\tilde{r}_t = \alpha_{t-1} + \beta_{t-1}\gamma_{t-1} + \beta_{t-1}v_t + \tilde{u}_t,$$

Mx1 vectors

- Too many degrees of freedom, need additional restrictions:

$$\beta_{t-1} = b_{t-1}\beta,$$

MxN matrix of observable characteristics NxK vector of parameters

$$\tilde{r}_t = b_{t-1}\tilde{f}_t + \tilde{\varepsilon}_t,$$

$\tilde{f}_t := \beta(\gamma_{t-1} + v_t)$ is a new $N \times 1$ vector of latent factors, and $\tilde{\varepsilon}_t := \alpha_{t-1} + \tilde{u}_t$.

- → Barra's model
  - Include several dozens of characteristics and is heavily overparameterized→

- Instrumented PCA (IPCA) – Kelly et al.(2019)
  - Inherits Barra's versatility and tractability, yet avoids its statistical inefficiency via a built-in dimension reduction:

$$\tilde{r}_t = b_{t-1}\beta f_t + \tilde{\varepsilon}_t,$$

β and $\{f_t\}$ have $N \times K$ and $K \times T$ unknown parameters, respectively.

  - IPCA employ a linear approximation for risk exposure based on observable characteristics data.
  - Nonlinearity→
- Conditional autoencoder model – Gu et al.(2021)
  - Replace the linear beta specification with a more flexible beta function

Methodologies

- **High-dimensional statistical methods** are increasingly relevant for empirical asset pricing analysis.

  - Low dimension → high dimension
  - Few assets → individual stocks and so on
  - Few factors → factor zoo
  - Classical methods → machine learning & deep learning

Measuring expected returns

- Return prediction is critical to developing a clearer understanding of financial markets
- Three literature strands:
  - Cross-sectional: $r_{i,t+1} = f(X_{i,t})$ X: a small list of stock-level characteristic
  - Time-serial: $r_{t+1} = f(X_t)$ X: a small set of predictors
    - Challenge: large number of predictors
  - Newly emerging machine learning methods
    - Delves little into understanding the economic mechanisms

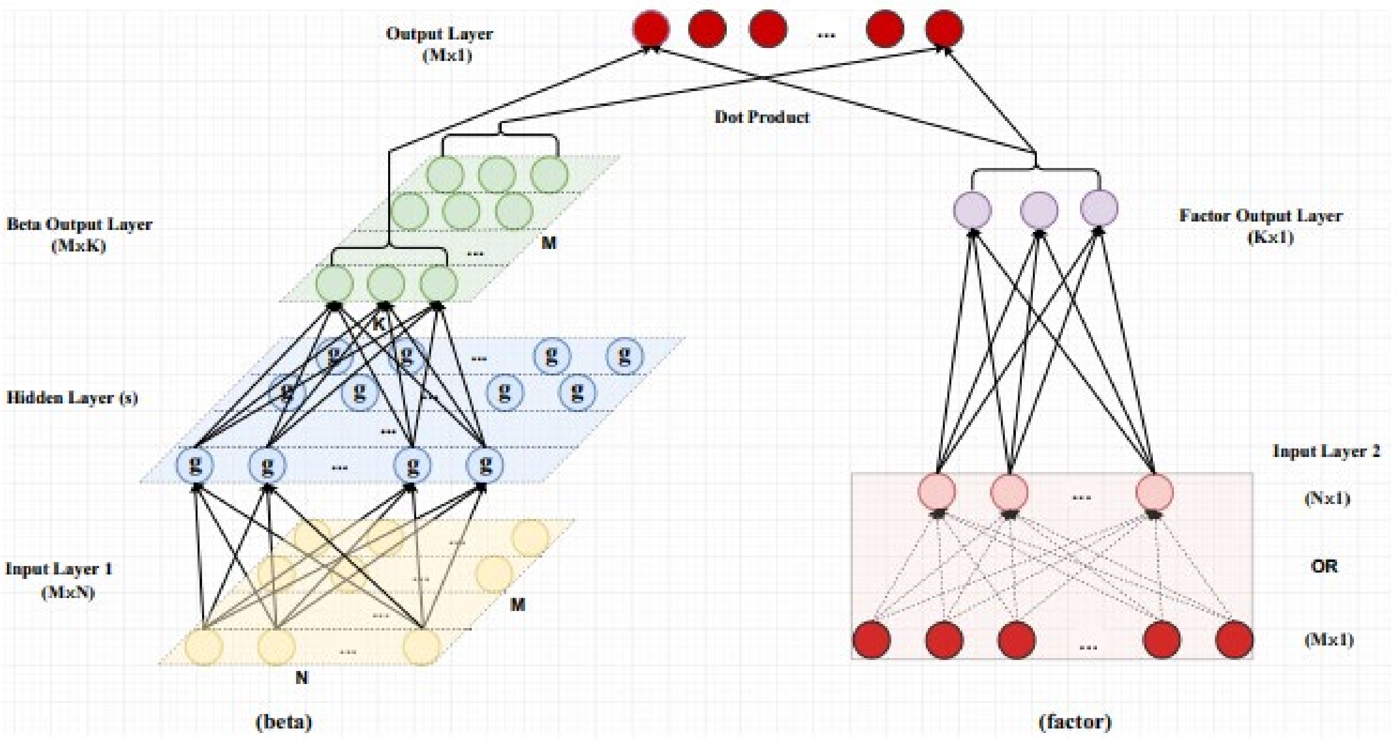
Estimating factors and exposures

- Variance of assets = systematic risk + idiosyncratic risk
- Different modeling strategies:
  - Whether factors and their exposures are known?
  - Whether models use a conditional or unconditional risk decomposition?

- 1. TSR and CSR
  - If factors are known, run TSR for each asset → β
  - If factors are latent, exposures are observable, run CSR at each t → β
- 2. PCA
  - If neither factors nor loadings are known, run PCA to extract latent factors
  - Assumptions:
    - The covariance matrix of assets features a few dominant factors that drive most covariance
    - Sentiment investors don't have too much influence.
  - Singular value decomposition(SDV)
  - Difficult to interpret

- 3. risk premia PCA
  - The PCA above assumes $\alpha = 0$
  - And this take α into consideration
- 4. instrumented PCA
  - Time-varying (conditional)

    $\min_{\beta,\{f_t\}} \sum_{t=2}^{T} \|\tilde{r}_t - b_{t-1}\beta f_t\|^2.$



- 5. autoencoder learning
  - Left: factor loadings are nonlinear function of covariates
  - Right: factors as portfolios of individual stock returns

- PCA→IPCA→Autoencoder

- Several generic algorithms in deep learning:
    - Training, validation and testing
    - Regularization techniques

$$\mathcal{L}(\theta;\cdot) = \frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\left\|\tilde{r}_{i,t} - \beta'_{i,t-1}f_t\right\|^2 + \phi(\theta;\cdot)$$

    - Optimization algorithms
- Matrix completion
    - Adam; batch normalization……

Estimating risk premia

- Investors should be compensate for their exposure to factors
- Tradable factors: sample average return of the factor
- Untradable factors: eg. Consumption, inflation, liquidity…
  - 1. classical two-pass regression
    - Requires all factors observable
    - Two steps:
      - TS regression to estimate β
      - CS regression of average returns($\bar{r}$) on the $\hat{\beta}$ to estimate risk premia
    - Can replace OLS with GLS(generalized LS), but no asymptotic efficiency gain

- 2. factor mimicking portfolio
  - Two approaches:
    - 1. Fama-MacBeth regression:
      - TS regression to estimate β
      - Regress realized returns at each time t onto $\hat{\beta}$ to estimate $\widehat{\gamma_t}$ $\qquad \widehat{\gamma}_t = (\widehat{\beta}^\intercal \widehat{\beta})^{-1} \widehat{\beta}^\intercal r_t$
        - $\hat{\gamma}$ itself is a portfolio return, then estimate the risk premium as the time series average of $\widehat{\gamma_t}$
    - 2. Maximal-correlation factor-mimicking approach
      - Project factor onto a set of basis assets, which yields weights of the mimicking portfolio

$$w_g = \mathrm{Var}(y_t)^{-1}\mathrm{Cov}(y_t, g_t),$$

    - Use the same factors in the above two approaches and they are equivalent
    - The second approach is better: don't require a fully specified factor model
    - $\rightarrow$ curse of dimensionality

- 3. three-pass regressions
  - Cope with high-dimension problem
  - Three steps:

1. The first-pass is an SVD of $\bar{R}$ to obtain $\widehat{\beta}$ and $\widehat{V}$ as in 13.

2. The second pass runs a cross-sectional OLS, 24, to obtain risk premia of $\widehat{V}$.

3. Finally, the third pass projects g_t onto $\widehat{V}$:

$$\widehat{\eta} = \bar{G}\widehat{V}^{\intercal}(\widehat{V}\widehat{V}^{\intercal})^{-1},$$

thus recovering the weights of the mimicking portfolio.

- Weak factors
  - An issue of two-pass regression: weak identification→
    - Useless factors
    - Small beta
    - Factor collinearity
  - Penalized two-pass regression/ IV estimator to correct bias…
  - Giglio et al(2021): active test asset selection; iterative supervised PCA →
  - Which test assets?
    - Tradable factors: independent of the test assets
    - Non-tradable factors: very important
    - 1. standard set of portfolios sorted on a few characteristics
    - 2. portfolios sorted on a much larger set of characteristics
    - 3. a specific factor of interest
      - Commonly used: Estimate stock-level betas on a given factors, then sort assets into portfolios based on the estimated exposure.

Estimating the SDF and its loadings

- Factor risk premium and SDF
  - $E(\tilde{m}\tilde{R}) = 1$, $1 = cov(\tilde{m}, \tilde{R}) + E(\tilde{m})\, E(\tilde{R})$
  - $\rightarrow E(\tilde{R}) - R_f = -R_f cov(\tilde{m}, \tilde{R})$
  - $\rightarrow \ m_t = 1 - b^\intercal v_t,$ where $b = \Sigma_v^{-1}\gamma$ and Σ_v is the covariance matrix of factor innovations.
  - SDF loading b and risk premia γ are directly related through the covariance matrix of the factor, but they differ in interpretation:
    - SDF loading: whether that factor is useful in pricing the cross section of returns
    - A factor could have nonzero risk premium without appearing in the SDF
    - Makes sense to tame factor zoo by testing SDF loadings instead of risk premium

- Estimate SDF loadings
  - generalized method of moment

$$\mathrm{E}(m_t r_t) = 0_{N\times 1}, \quad \mathrm{E}(v_t) = 0_{K\times 1}.$$

Factor innovation

Since there are in total $K + N$ moments with $2K$ parameters (μ and b) in general, we need $N \geq K$ to ensure the system is identified.

$f_t = \mu + v_t, \quad b = \Sigma_v^{-1}\gamma$

The GMM estimator is thereby defined as the solution to the optimization problem:

$$\min_{b,\mu} \widehat{g}_T(b,\mu)^\intercal \widehat{W} \widehat{g}_T(b,\mu), \tag{30}$$

- PCA-based methods
  - The SDF can be represented as a function of a few dominant sources of return variation

$$\widehat{m}_t = 1 - \widehat{\gamma}^{\intercal} \widehat{v}_t,$$

- Penalized regressions
  - Represent SDF in terms of a set of tradable test asset returns

$$\underline{m}_t = 1 - \underline{b}^{\intercal}(r_t - \mathrm{E}(r_t))$$

- Double machine learning

In the spirit of DML, Feng et al. (2020) select controls from $\{\widehat{C}_h\}$ via two respective lasso regressions: $\bar{r}$ onto $\widehat{C}_h$ and $\widehat{C}_g$ onto $\widehat{C}_h$. The selected controls, denoted by $\widehat{C}_{h[I]}$, along with $\widehat{C}_g$, serve as regressors in another cross-sectional regression of $\bar{r}$. The resulting estimator of b_g,

$$\widehat{b}_g = (\widehat{C}_g^{\intercal} \mathbb{M}_{\widehat{C}_{h[I]}} \widehat{C}_g)^{-1} (\widehat{C}_g^{\intercal} \mathbb{M}_{\widehat{C}_{h[I]}} \bar{r}),$$

- Parametric portfolios and deep learning SDFs

directly parametrizing portfolio weights as functions of asset characteristics, then estimate the parameters by solving a utility optimization problem:

$$\max_{\theta} \frac{1}{T} \sum_{t=2}^{T} \mathrm{U}\left(\sum_{i=1}^{N_t} w(\theta, b_{i,t-1}) \tilde{r}_{i,t} \right),$$

Model specification tests and model comparison

- GRS test and extensions
  - Focus on $\alpha = 0$: if the factor model reflects the true SDF, then it should price all test assets with zero alpha

$$\mathbb{H}_0 : \alpha_1 = \alpha_2 = \ldots = \alpha_N = 0.$$

  - Limitation: it requires that T>N+K $\rightarrow$
    - Pesaran and Yamagata(2017): a simple quadratic test
    - Fan et al.(2015): impose a sparsity structure on covariance matrix
- Model comparison tests
  - The classical GRS test: whether the factors achieve the maximal Sharpe ratio
  - Others:
    - Barillas and Shanken(2017): the ability to price all returns
    - Barillas et al.(2020): compare Sharpe ratio
    - ……

- Bayesian approach
  - As the set of candidate models expands, model comparison via pairwise comparison becomes a tough task. →
  - Barrillas and Shanken(2018): a Bayesian procedure that computes model probabilities for a collection of asset pricing models with tradable factors.

Alphas and multiple testing

- Alpha: the portion of expected returns that can’t be explained by risk exposures.
- Anomaly: a portfolio with significant alpha
- Data-snooping concern and MT issue
  - Replace multitude null hypothesis with one single null hypothesis
  - Control false discovery (sacrifice power)

  - Widen confidence internals and raise p-values, but do not alter the underly point estimate

Conclusion

- ML is neither an empirical panacea nor a substitute for economic theory and the structure it lends to empirical work.
- The most promising direction for future empirical asset pricing research is developing a genuine fusion of economic theory and ML.
- ML factor models are one such example of this fusion.