Bayesian Solutions for the Factor Zoo: We Just Ran Two Quadrillion Models
by Bryzgalova, Huang, Julliard

Harjoat S. Bhamra

Imperial College Business School

2021

Motivation I

- Major aim of factor models is to understand
  - what drives expected risk premia for large number of assets (N_A) with a small number of factors (N_F), or
  - which factors drive unexpected changes in SDF

$$\boxed{N_F \ll N_A} \tag{1}$$

- Rate of factor production far exceeds number of assets

$$\boxed{\frac{dN_F}{dt} \gg \frac{dN_A}{dt}} \tag{2}$$

Motivation II

- Understanding individual papers
- A single paper focuses on dimensionality reduction.
- But each paper tries to find new factors!
- Some factors are spurious — easy to identify an irrelevant factor as important
  - Gospodinov, Kan, and Robotti (2019)

We have a factor selection problem.

Motivation III

- With many papers, we have over 100 factors to choose from – a Factor Zoo
  - Cochrane (2011), Harvey, Liu, and Zhu (2015), McLean and Pontiff (2016), & Hou, Xue, and Zhang (2017)
- Hard to decide which models are important

We have a model selection problem.

This Paper

- Addresses spurious factor and model selection (very important)
- Develops a Bayesian version of Fama-MacBeth
- Not as simple is it sounds!
- This paper does the hard work for you. Be grateful!

Dealing with Spurious Factors

- Fama-MacBeth
  - index factors by $k \in \{1, \ldots, K\}$, firms by $n \in \{1, \ldots, N\}$, time by $t \in \{1, \ldots, T\}$
  - $K \ll N$, want T large

$$N \text{ times series regressions: } R_{n,t} = \alpha_n + \sum_{k=1}^{K} \beta_{n,k} f_{k,t} + \epsilon_t \tag{3}$$

obtain estimates: $\hat{\beta}_{n,k}$, K for each firm n

$$\text{cross-sectional regression: } \overline{R}_n = \lambda_c + \alpha_n + \sum_{k=1}^{K} \lambda_k \hat{\beta}_{n,k} \tag{4}$$

$$\hat{\boldsymbol{\lambda}} = (\beta^\top \beta)^{-1} \beta^\top \overline{\boldsymbol{R}} \tag{5}$$

- If $\beta_{n,k} = \frac{C_n}{\sqrt{T}}$, i.e. factor k is spurious, then $\hat{\lambda}_k \nrightarrow 0$ as $T \to \infty$

Bayesian Fama-MacBeth: Time Series

- Assume errors are normally distributed, Σ is variance-covariance matrix of errors.
- probability distributions replace estimates
- Define matrix B

$$B = \begin{pmatrix} \alpha_1 & \beta_{11} & \cdots & \beta_{1K} \\ \vdots & \vdots & \ldots & \vdots \\ \alpha_N & \beta_{N1} & \cdots & \beta_{NK} \end{pmatrix} \tag{6}$$

- Bayes' Law for times series regressions: prior $\rightarrow$ posterior

$$\underbrace{p(B, \Sigma | \text{data})}_{\text{posterior}} \propto \underbrace{p(\text{data} | B, \Sigma)}_{\text{likelihood}} \underbrace{p(B, \Sigma)}_{\text{prior}} \tag{7}$$

- Jeffrey's prior $p(B, \Sigma) \propto [\det(\Sigma)]^{-\frac{N+1}{2}}$

Bayesian Fama-MacBeth: Cross - Section

- Assume $\alpha_i \sim N[0, \sigma^2]$
- Bayes' Law for cross-sectional regression: prior $\rightarrow$ posterior

$$\underbrace{p(\lambda, \sigma^2 | \text{data}, B, \Sigma)}_{\text{posterior}} \propto \underbrace{p(\text{data} | \lambda, \sigma^2)}_{\text{likelihood}} \underbrace{\pi(\lambda, \sigma^2)}_{\text{prior}} \tag{8}$$

- Jeffrey's prior $\pi(\lambda, \sigma^2) \propto \sigma^{-2}$
- The posterior of λ conditional on B, Σ and the data, is a Dirac distribution at $\left(\beta^\top \beta\right)^{-1} \quad \underbrace{\beta^\top \boldsymbol{\alpha}}_{\text{switches sign}}$
  - sign switching means λ can be zero, avoiding spurious factor problem

Model Selection

- Index two models by γ and model γ'
- Bayes' Theorem with equal prior probability to each model gives posterior probability of model γ as

$$\Pr(\boldsymbol{\gamma} \mid \text{ data }) = \frac{p(\text{ data } \mid \boldsymbol{\gamma})}{p(\text{ data } \mid \boldsymbol{\gamma}) + p(\text{ data } \mid \boldsymbol{\gamma}')}$$

- Flat prior for λ make marginal likelihood $p(\text{ data } \mid \gamma)$ blow up if there is a spurious factor in model γ
- Use spike and slab prior – allow for possibility a factor is irrelevant in prior (spike)

Interpreting spike and slab prior

- We know that the Sharpe ratio on the market is around 0.4 in annual units. We also also know factor Sharpe ratios. This is built into the prior.

$$D^{KL}[\mathbb{Q}|\mathbb{P}] = -\frac{1}{dt} E_t \left[d \ln \frac{\Lambda_{t+dt}}{\Lambda_t} \right] \approx \frac{1}{2} 0.4^2 = \frac{1}{2} \sum_{1 \leq i,j \leq K} \Theta_{i,t} \Theta_{j,t} \rho_{ij,t} \qquad (9)$$

  - Θ_i is the Sharpe ratio for factor i, ρ_{ij} is the correlation between factors i and j
  - If we constrain $D^{KL}[\mathbb{Q}|\mathbb{P}]$ to be close to $\frac{1}{2}0.4^2$, we are constraining ourselves to be close to a hyperellipsoid in K-dimensional factor space.
  - If we constrain Θ_i to lie in a small interval and fix the correlations, we are constraining ourselves to hypercubes in K-dimensional factor space, which are centred on the above hyperellipsoid,
  - Would be nice to go from constraint to prior formally.

Questions: Priors

- Much of the power of Bayesian methods comes from the selection of a prior guided by theory
- Why not more weight in the prior on factors which come from theoretical models? We are in 2021!
- Why not assume non traded factors are irrelevant?

Are linear factors model very useful?

- It may not be possible to approximate the SDF well via small set of linear factors – this paper says
  'the "true" latent SDF is dense in the space of observable factors, that is, a large subset of the factors proposed in the literature is needed to fully capture its pricing implications'
- Hard to pick the linear factor model to use.
- We need factor models that can deal with

$$\frac{d\Lambda_t}{\Lambda_t} - E_t\left[\frac{d\Lambda_t}{\Lambda_t}\right] = -\sum_{i=1}^{K}\Theta_{i,t}dZ_{i,t}, \tag{10}$$

  where risk prices Θ_i are stochastic and nonlinear – means importance of factors will be time varying – look at HML in recent data...
- Dittmar (2002)
  'Our test results indicate that preference-restricted nonlinear pricing kernels are both admissible for the cross section of returns and are able to significantly improve upon linear single- and multifactor kernels. Further, the nonlinearities in the pricing kernel drive out the importance of the factors in the linear multi-factor model.'
- How about a Bayesian nonlinear factor model?– using priors in place of restrictions to particular preferences.

Summary

- Very clean and impressive methodology.
- Could be used in further work with priors informed by theory.
- Maybe the Bayesian approach is telling us something about how useful (or not) linear factor models are.