Discussion on

A penalized two-pass regression to predict stock returns with time-varying risk premia

by Bakalli, Guerrier, and Scaillet

Marcelo C. Medeiros (PUC-Rio)

ABFR Webinar

October 27, 2022

Overview

- Penalized two-pass regression with time-varying factor loadings:

  1. First pass enforces sparsity for the time-variation drivers + no-arbitrage restrictions
  2. The second pass delivers risk premia estimates to predict equity excess returns

- Monte Carlo and empirical results corroborate the method

The Structural Model

1. Factor model with time-varying intercepts and loadings

$$R_{it} = a_{it} + \boldsymbol{b}'_{it}\boldsymbol{f}_t + \varepsilon_{it}$$

- $\boldsymbol{f}_t$ takes values on $\mathbb{R}^K$
- $\mathbb{E}(\varepsilon_{it}|\mathcal{F}_{t-1}) = 0 \longrightarrow$ linearity is assumed
- $\mathbb{C}(\varepsilon_{it}, f_{kt}|\mathcal{F}_{t-1}) = 0$
- Asset pricing restriction: $a_{it} = \boldsymbol{b}'_{it}\boldsymbol{\nu}_t$, $\boldsymbol{\nu}_t$ unique and $\mathcal{F}_{t-1}-$measurable $\longrightarrow \mathbb{E}(R_{it}|\mathcal{F}_{t-1}) = \boldsymbol{b}'_{it}\boldsymbol{\lambda}_t$, $\boldsymbol{\lambda}_t = \boldsymbol{\nu}_t + \mathbb{E}(\boldsymbol{f}_t|\mathcal{F}_{t-1})$

2. Sparse time-varying factor loadings

$$\boldsymbol{b}_{it} = \boldsymbol{A}_i + \boldsymbol{B}_i \boldsymbol{Z}_{t-1} + \boldsymbol{C}_i \boldsymbol{Z}_{it-1}$$

- $\boldsymbol{A}_i \in \mathbb{R}^K$ does not contain null elements
- $\boldsymbol{B}_i \in \mathbb{R}^{K \times p}$ and $\boldsymbol{C}_i \in \mathbb{R}^{K \times q}$ are sparse matrices of coefficients
- $\boldsymbol{Z}_{t-1}$ takes values on $\mathbb{R}^p$ and is a vector of common instruments
- $\boldsymbol{Z}_{it-1}$ takes values on $\mathbb{R}^q$ and is a vector of firm-specific characteristics

3. Sparse time-varying risk premia

$$\boldsymbol{\lambda}_t = \boldsymbol{\Lambda}_0 + \boldsymbol{\Lambda}_1 \boldsymbol{Z}_{t-1}$$
$$\mathbb{E}(\boldsymbol{f}_t | \mathcal{F}_{t-1}) = \boldsymbol{F}_0 + \boldsymbol{F}_1 \boldsymbol{Z}_{t-1}$$

- $\boldsymbol{\Lambda}_0 \in \mathbb{R}^K$ and $\boldsymbol{F}_0 \in \mathbb{R}^K$ do not contain null-elements
- $\boldsymbol{\Lambda}_1 \in \mathbb{R}^{K \times p}$ and $\boldsymbol{F}_1 \in \mathbb{R}^{K \times p}$ are sparse matrices of coefficients

1. The setup of the paper is low-dimensional in the sense that K, p and q are assumed to be known and fixed. This should be made very clear in the very beginning of the paper.

2. Therefore, why do we need sparsity?

The Structural Model

Some Comments (2/2)

1. Is the martingale difference assumption on ε_{it} (linearity) necessary? Why not just assuming that the errors are uncorrelated?

   – Gu, Kelly and Xiu (RFS, 2020) found compelling evidence of nonlinearities in asset pricing models.

2. Under model assumptions all time dependence on the returns is given by the factor dynamics and the time-varying structure of the intercept and the loadings. The authors may wish to discuss a bit about this in the paper.

The Reduced Form

Putting things together...

$$
\begin{aligned}
R_{it} = {} & \underbrace{\boldsymbol{A}_i'(\boldsymbol{\Lambda}_0 - \boldsymbol{F}_0) + \boldsymbol{A}_i'\boldsymbol{f}_t}_{\text{time-invaryinig component}} \\
& + \boldsymbol{A}_i'(\boldsymbol{\Lambda}_1 - \boldsymbol{F}_1)\boldsymbol{Z}_{t-1} + \boldsymbol{Z}_{t-1}'\boldsymbol{B}_i'(\boldsymbol{\Lambda}_0 - \boldsymbol{F}_0) \\
& + \boldsymbol{Z}_{it-1}'\boldsymbol{C}_i'(\boldsymbol{\Lambda}_0 - \boldsymbol{F}_0) \\
& + \boldsymbol{Z}_{t-1}'\boldsymbol{B}_i'(\boldsymbol{\Lambda}_1 - \boldsymbol{F}_1)\boldsymbol{Z}_{t-1} + \boldsymbol{Z}_{it-1}'\boldsymbol{C}_i'(\boldsymbol{\Lambda}_1 - \boldsymbol{F}_1)\boldsymbol{Z}_{t-1} \\
& + \boldsymbol{Z}_{t-1}'\boldsymbol{B}_i'\boldsymbol{f}_t + \boldsymbol{Z}_{it-1}'\boldsymbol{C}_i'\boldsymbol{f}_t + \varepsilon_{it}
\end{aligned}
$$

The Reduced Form

$$R_{it} = \boldsymbol{\beta}_i' \boldsymbol{x}_{it} + \varepsilon_{it}$$

- $\boldsymbol{x}_{it}$ is a nonlinear function of $\boldsymbol{f}_t$, $\boldsymbol{Z}_{t-1}$, and $\boldsymbol{Z}_{it-1}$
- Model is linear in the parameters but nonlinear in the original variables
- no-arbitrage assumption imposes restrictions on $\boldsymbol{\beta}_i$

- Estimate $\boldsymbol{\beta}_i$ in

$$R_{it} = \boldsymbol{\beta}_i' \boldsymbol{x}_{it} + \varepsilon_{it}$$

  by
  - "unrestricted" adaptive LASSO (aLASSO): no-arbitrage is not imposed
  - adaptive latent group LASSO (aOGL): no-arbitrage is imposed
- Only time-varying terms are shrunk $\longrightarrow$ shrinkage towards the constant model
- Result: under some high-level assumptions, the aOGL estimator converges to zero for the truly zero parameters and to a Gaussian random variable for the remaining ones

1. High-level assumptions on $\boldsymbol{x}_{it}$. I miss more primitive assumptions on $\boldsymbol{Z}_{t-1}$ and $\boldsymbol{Z}_{it-1}$.
   - For instance what are the moment structure of these random vectors?
   - Are they time-dependent? α-mixing, for example?
   - For example, Assumption B.1 bounds $\boldsymbol{x}_{it}$. Is this expected given the nature of the data in empirical applications?
   - In empirical applications, $\boldsymbol{Z}_{t-1}$ and $\boldsymbol{Z}_{it-1}$ can be highly persistent and non-Gaussian. For example, inflation or volatility
   - Unbalanced regressions

2. The role of sparsity
   - Sparsity does not seem necessary to derive the result. Is this necessary for the no-arbitrage restriction?
   - What does happen if a ridge penalty is used instead?
3. Discuss more the role of the initial estimator
4. Although model selection consistency is not derived in the paper, this could be a nice additional result
5. It is difficult to follow the proofs without having the assumptions in GOS. You should state them at least in the Appendix

- Smart estimation to recover $\boldsymbol{\nu}$
- Estimation of the factor dynamics by aLASSO
- Result: the estimator for $\boldsymbol{\nu}$ is consistent under some high-level conditions and both n, T diverging
- Result: the estimator for $\boldsymbol{\Lambda}$ is consistent under some high-level conditions and both n, T diverging

1. Sparsity again: why do we need sparsity on the factor dynamics?
2. Why aLASSO instead of just LASSO? It seems that correct variable selection is not an issue here, right?
3. It is not clear if the penalty parameter is the same for all estimations or not

Some Comments on the Simulations

1. Is the DGP too restrictive? Relax the normality assumption of the errors
2. Both the aLASSO and the aOLG methods strongly under-select the relevant variables. This looks very strange to me. Smallish coefficients?
3. Related to the previous comments, it would be nice to get an idea of the order of magnitude of the nonzero coefficients. Are the RMSEs large or not?
4. The aLASSO should be also consistent. Why is the performance so "bad"? Small samples?
5. The MAPEs in Table 3 seem quite similar. Any comments?

Some Comments on the Simulations

1. Very different relative performance depending on the sample considered. Structural breaks or just the size of the estimation period?
2. The $\lambda_t \& \nu$ model seems very competitive specially during the first subsample. Structural breaks or just the size of the estimation period?
3. What happens if nonlinear ML models as in Gu, Kelly, and Xiu (RFS, 2020) are considered?

Some General Comments

1. The paper will benefit from having a guide to empirical implementation.
2. Positioning the paper wrt the recent literature on factor zoo would be interesting
3. Also, it would be nice to compare with Projected/Instrumented PCA (factors are not observed)
4. Idea: Combine the results in Fan, Masini, and Medeiros (2021) with the no-arbitrage restrictions

5. And finally,

THIS IS A GREAT PAPER!

CONGRATULATIONS!

VERY Minor Comments

- In page 4, paragraph after equation (1), Z_{t-1} is not defined
- Assumption A.1: "sparse matrices of coefficients" and not "sparse matrices of coefficient"
- Typo in equation (9): δ_g should be inside the summation