Extracting Statistical Factors When Betas Are Time-Varying

Patrick Gagliardini, Hao Ma
Università della Svizzera italiana (USI, Lugano)
& Swiss Finance Institute (SFI)

Contact Information:
Institute of Finance
Università della Svizzera italiana
Via Giuseppe Buffi 13, 6900 Lugano, Switzerland

Email: hao.ma@usi.ch

Abstract

This paper deals with identification and inference on the unobservable *conditional* factor space and its dimension in large unbalanced panels of asset returns. The model specification is nonparametric regarding the way the loadings vary in time as functions of common shocks and individual characteristics. The number of active factors can also be time-varying as an effect of the changing macroeconomic environment. The method deploys Instrumental Variables (IV) which have full-rank covariation with the factor betas in the cross-section. It allows for a large dimension of the vector generating the conditioning information by machine learning techniques. In an empirical application, we infer the conditional factor space in the panel of monthly returns of individual stocks in the CRSP dataset between January 1971 and December 2017.

Introduction

Motivation

Asset Pricing literature on time-varying beta specifications has mostly focused on models with observable factors. However, there is a great latitude in the choice of observable economic factors. In this paper, we aim to find a *dynamic* PCA method for inference in *latent* factor models with *time-varying* betas.

Contributions

Methodological aspects

1. Deploy *No-Arbitrage* restrictions with conditioning information
2. (Almost) *"model-free"* regarding the dynamics of betas
3. *First* paper that allows for *time-varying* number of conditional factors and provide a *consistent* selection procedure
4. Cope with *large-dimensional* conditioning information via *machine learning* techniques
5. Use large *unbalanced* panels of *individual* stock returns
6. Develop *asymptotic theory* $(n, T \to \infty)$ building on results for plug-in Sieve estimation and Double Machine Learning.

Empirical findings

1. There are two dominant factors across the sample period, together explaining over 70% measured by the AEP ratio.
2. The number of factors tends to be smaller during recession periods.
3. The first conditional factor is best explained by *MKT*, and the second one is best spanned by *SMB*.

No-Arbitrage Conditional Factor Model

Consider a conditional factor model for individual asset i in period t:

$$y_{i,t} = a_{i,t-1} + b'_{i,t-1} f_t + \varepsilon_{i,t}, \quad t = 1, ..., T, \quad i = 1, ..., n, \tag{1}$$

- $y_{i,t}$ is the excess return,
- $a_{i,t-1}$ is the abnormal return, a $\mathcal{F}_{i,t-1}$-measurable scalar
- $b_{i,t-1}$ is the time-varying beta, a $\mathcal{F}_{i,t-1}$-measurable $k_t \times 1$ vector
- f_t is the systematic risk factor, a $\mathcal{F}_t$-measurable $k_t \times 1$ vector
- $\varepsilon_{i,t}$ is the idiosyncratic error term

No-arbitrage restriction (Gagliardini, Ossola and Scaillet (2016)):

$$a_{i,t-1} = b'_{i,t-1} \nu_{t-1}, \tag{2}$$

Insert (2) into (1) and get the no-arbitrage conditional factor model:

$$\boxed{y_{i,t} = b'_{i,t-1} g_t + \varepsilon_{i,t},} \tag{3}$$

where $g_t = \nu_{t-1} + f_t$ and $\lambda_t = \nu_{t-1} + E[f_t | \mathcal{F}_{t-1}]$ is the risk premium.

Identification Strategy

Assumption 1. *There exists a $K \times 1$ $(K \geq k)$ vector of instrumental variables $w_{i,t-1}$ measurable w.r.t. $\mathcal{F}_{i,t-1}$ such that:*

$$(i) \ \underset{n \to \infty}{\text{plim}} \frac{1}{n} \sum_{i=1}^{n} w_{i,t-1} \varepsilon_{i,t} = 0,$$

$$(ii) \ \underset{n \to \infty}{\text{plim}} \frac{1}{n} \sum_{i=1}^{n} w_{i,t-1} b'_{i,t-1} =: \Gamma_{t-1} \text{ is a } K \times k \text{ full-rank matrix, P-a.s.}$$

Instrument-Weighted Portfolio Returns ξ_t

Define the large n limit of cross-sectional portfolio returns:

$$\xi_t = \underset{n \to \infty}{\text{plim}} \frac{1}{n} \sum_{i=1}^{n} w_{i,t-1} y_{i,t}, \tag{4}$$

which is measurable w.r.t. $\mathcal{F}_t$. Then, (3) and Assumption 1 imply:

$$\boxed{\xi_t = \Gamma_{t-1} g_t} \tag{5}$$

Identification of Number of Factors k_t

The conditional variance of ξ_t is a $K \times K$ symmetric matrix:

$$V(\xi_t | \mathcal{F}_{t-1}) = \Gamma_{t-1} V(g_t | \mathcal{F}_{t-1}) \Gamma'_{t-1} \tag{6}$$

The number of latent conditional factors k_t is identifiable by the rank:

$$\boxed{k_t = Rank(V[\xi_t | \mathcal{F}_{t-1}])} \tag{7}$$

Identification of Latent Factors f_t

Let J_{t-1} be a $\mathcal{F}_{t-1}$-measurable $K \times k$ full-rank matrix whose columns are the standardized eigenvectors of $V(\xi_t | \mathcal{F}_{t-1})$ associated to non-zero eigenvalues.

Assumption 2. *Without loss of generality, we assume that the following normalization restriction holds for the latent factors:*

$$\Gamma_{t-1} = J_{t-1}, \forall t \tag{8}$$

Therefore, g_t is identifiable as:

$$\boxed{g_t = J'_{t-1} \xi_t,} \tag{9}$$

Assumption 3. *Without loss of generality, we assume $E[f_t | \mathcal{F}_{t-1}] = 0$.*

Under Assumption 2 and 3, the factor vector f_t is identifiable as:

$$\boxed{f_t = g_t - E[g_t | \mathcal{F}_{t-1}]} \tag{10}$$

Estimation Methodology

Assumption 4. *The information set $\mathcal{F}_t$ is generated by the observable d-dimensional vector Markov process Z_t.*

This assumption implies:

$$E[\zeta_t | \mathcal{F}_{t-1}] = E[\zeta_t | Z_{t-1}] =: \psi^{\zeta}(Z_{t-1})$$

for a function $\psi^{\zeta}(\cdot)$ and any random vector ζ_t. Since Z_t could be large-dimensional, machine learning techniques are used in estimating conditional expectations and variances.

- **Post-Lasso** method used in e.g. Belloni et al. (2012)
- **Artificial Neural Networks** with different network structures

Double Machine Learning Inference on Avg. Cond'l Features

Let the finite-dimensional parameter $c = c(\theta)$ be defined by

$$c = E[\gamma_t] = E[\varphi(\theta(Z_{t-1}))]$$

e.g. average conditional correlation.

Double Machine Learning (DML): use the "locally robust" moment restriction

$$E\left[\varphi(\theta(Z_{t-1})) - c + \alpha(Z_{t-1})'(\zeta_t - \theta(Z_{t-1}))\right] = 0 \tag{11}$$

where function $\alpha(\cdot)$ is the Rietz representer of the Gateaux derivative

$$\lim_{\tau \to 0} \frac{c(\theta_0 + \tau(\theta - \theta_0)) - c(\theta_0)}{\tau} = \langle \alpha, \theta - \theta_0 \rangle$$

Split the sample in subintervals I_ℓ, $\ell = 1, ..., L$, and the DML estimator of c is given by

$$\hat{c} = \frac{1}{T} \sum_{\ell=1}^{L} \sum_{t \in I_\ell} \left[\varphi(\hat{\theta}_\ell(Z_{t-1})) + \hat{\alpha}_\ell(Z_{t-1})'(\hat{\zeta}_t - \hat{\theta}_\ell(Z_{t-1}))\right]$$

where $\hat{\theta}_\ell$ and $\hat{\alpha}_\ell$ are obtained using observations *not* in I_ℓ.

If $\|\hat{\theta}_l - \theta_0\| = o_p(T^{-\frac{1}{4}})$, then $\sqrt{T}(\hat{c} - c) \overset{d}{\to} N(0, \sigma^2)$

Empirical Results

U.S. equities panel data from CRSP and COMPUSTAT during period January 1971 - December 2017.

- $y_{i,t}$: monthly excess returns of individual stocks
- $w_{i,t}$: 15 firm characteristics from Freyberger, Neuhierl and Weber (2017)
- $\hat{\xi}_t$: 15 characteristics-based cross-sectional averages plus *Fama-French* 5 factors, *Momentum*, and *Betting Against Beta*
- Z_t : 19 variables - financial indicators from Goyal and Welch (2008), risk factors and macroeconomic variables

Accumulative Explanatory Power Ratio (AEP)

We use AEP ratio defined below to determine the number of time-varying factors k_t:

$$\hat{R}_{r,\tau} = \frac{1}{6} \sum_{t \in \tau} \frac{\sum_{j=1}^{r} \sigma_j[\hat{V}(\hat{\xi}_t | \mathcal{F}_{t-1})]}{Tr[\hat{V}(\hat{\xi}_t | \mathcal{F}_{t-1})]}, r = 1, 2, 3, 4$$

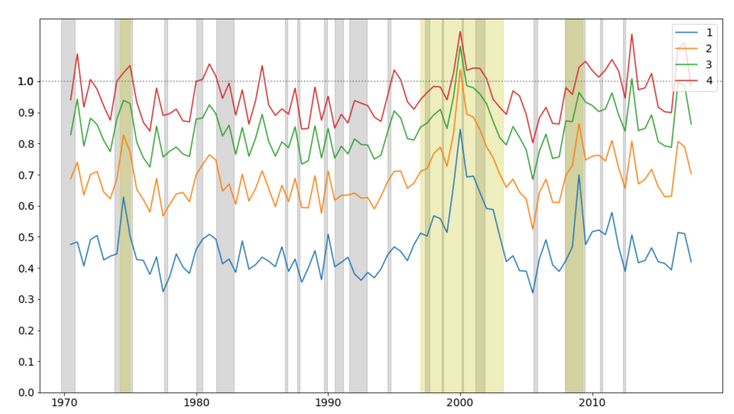

Figure 1: Accumulative Explanatory Power Ratio (6-month average)

The number of conditional factors is rather small. There are two dominant factors together with over 70% explanatory power across the whole sample period. Moreover, there tend to be fewer factors during recession periods.

Average Conditional Correlation

Figure 2 shows the average conditional correlation between our conditional latent factors $\hat{f}_t$ and each variable in the information set Z_t.

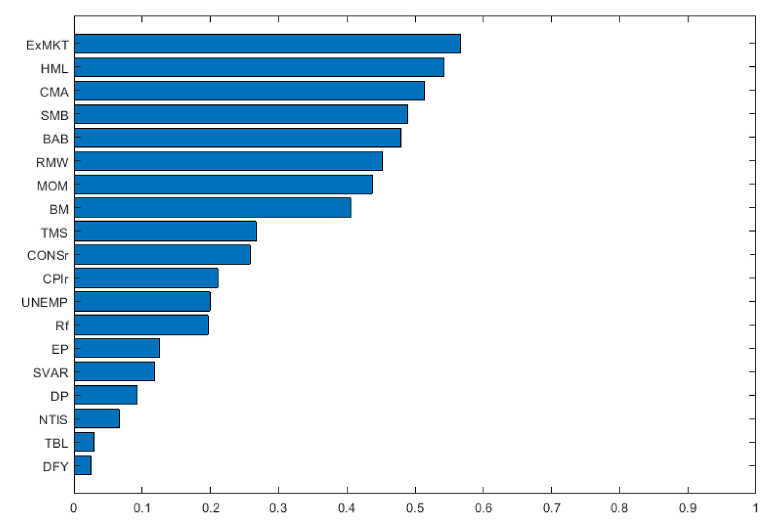

Figure 2: Estimation of $E[Corr_{t-1}(f_{1,t}, Z_t)]$

The above figure shows that the first conditional factor correlates most with the *Market* factor, at around 57%. According to another figure not shown here, the second conditional factor correlates most with the *Small-Minus-Big* factor, at about 48%.