5. Recent advances in cross-sectional asset pricing

Contents

1 Learning objective:

- **Reading list:** Giglio and Xui (2020) - Three Pass, Feng, Giglio, and Ziu (2020) - Taming the Factor Zoo
- Discuss and implement selected machine learnings methods for making valid inference in the presence of omitted factors, compare the findings to standard approaches, and reflect on the implication for asset pricing
- Using selected machine learnings methods to navigate the "factor zoo" and making valid inference in the presence of omitted factors

2 Suggested presentation 10 min

1. Empirical

Part I
Short version

3 Recent advances in cross sectional asset pricing

3.1 Introduction

- Investigate whether a risk factor is priced in the financial markets and estimates its risk premium
- Already know how to improve FMB to handle Errors in Variables and use GMM to estimate without EIV and make robust to autocorrelation without assuming normality
- Further problems from omitted variable bias or measurement errors
- Hundreds of potential risk factors out there, hard to include all as control
- Giglio and Xiu (2021) provide us with methodology to solve some of this problem
- Feng, Giglio and Ziu (2020) provide a method to evaluate the marginal contribution of a factor and tame the factor zoo

3.2 Measurement errors

3.2.1 Omitted variable bias

With true model:

$$Y = \beta X + \delta Z + E$$

Estimation with only X gives bias:

$$\mathbb{E}\left[\hat{\beta}|X\right] = \beta + (X'X)^{-1}\mathbb{E}\left[X'Z|X\right]\delta$$

This depends on the covariance between X and Z, as well as the association between Z and Y that we leave out.

In Fama-MacBeth, this gives us trouble because we bias the first-pass regression from not including Z, thus giving us biased $\hat{\beta}_1$.
In the second path regression we use biased $\hat{\beta}_1$ for estimation, and even leave out $\hat{\beta}_2$, giving us two sources of OVB bias for the risk premium.

Omitting important base assets will also spell trouble for a factor-mimicking portfolio approach.

3.2.2 Measurement error bias

If we only observe some factor subject to an error, such as is common for non-traded factors, this causes an attenuation bias in $\hat{\beta}_1$, leading to downward bias in $\hat{\gamma}$, and also biases the factor mimicking portfolio.

3.3 Three-pass methodology

A recent paper by Giglio and Xiu give us a method to handle both omitted variable bias and measurement error. We have the following model:

$$g_t = \delta + \eta v_t + z_t$$

where we want to estimate the risk factors for g_t, but we cannot observe all the true factors.
The risk premium that we want to estimate is the expected excess return of a portoflio with $\beta = 1$ w.r.t. g_t and $\beta = 0$ w.r.t. all other factors, implying that:

$$\gamma_g = \eta\gamma$$

Now we need some way to either estimate this product, or it's components, but we don't observe these constituents directly.
But, observing just any rotatin of the true factors allows us to get this risk premia through their three pass procedure!

1. Observe $\hat{v}_t = Hv_t$
2. Estimate $\hat{\beta}$ by regressing r_t^e onto $\hat{v}_t$
3. Estimate $\hat{\gamma}$ by regressing r_t^e onto $\hat{\beta}$
4. Estimate $\hat{\eta}$ by regressing g_t onto $\hat{v}_t$
5. Now we have $\hat{\eta}\hat{\gamma} = \eta H^{-1} H\gamma = \eta\gamma = \gamma_g$

3.3.1 PCA

This method identifies factors up to a rotation, providing us with $\hat{v}_t = Hv_t$ rather than v_t, as any rotation of the PCA components would explain the same amoutn of variance of the original data.

3.3.2 Three-pass steps:

1. PCA on full set of returns $\Rightarrow \hat{v}_t$
2. FMB: $\hat{\beta}$ from r_t^e onto $\hat{v}_t \Rightarrow \hat{\gamma}$ from r_t^e onto $\hat{\beta}$
3. Time series regression of g_t onto $\hat{v}_t$ to get $\hat{\eta}$

All we need to estimate $\hat{\gamma}_g$.
Benefits:

- No omitted variable bias
- No measurement error effects
- No need to include all relevant factors as in standard FMB
- Offers convergence between FMB and factor-mimicking

3.4 Asymptotic theory

As long as we consistently estimate the correct number of latent factors, the three-pass procedure does give us the right estimates.
As such we can conduct standard hypothesis tests on the significance of risk premias via normal t-tes.

3.5 Weak factors

$$g_t = \delta + \eta v_t + z_t$$

If $\eta = 0$, then we have weak factors, and the factor loadings or risk exposures would be poorly estimated.
Obviously, we want to test this null hypothesis of $\mathbb{H}_0 : \eta = 0$ vs. $\mathbb{H}_1 : \eta \neq 0$, which we can use a conventional Wald test for:
Here we use R^2 as measure of strength:

$$\hat{R}_g^2 = \frac{\hat{\eta}\hat{V}\hat{V}'\hat{\eta}'}{\overline{GG}'} \xrightarrow{p} R_g^2$$

where R_g^2 happens to be the true version from regression using v_t, not $\hat{v}_t$
Naturally hinges upon the assumption that latent factors are pervase, otherwise this test wouldn't mean anything.

Part II
Long version

4 Recent advances in cross-sectional asset pricing

4.1 Introduction

- A fundamental research questions in empirical asset pricing is whether a given risk factor is priced in the financial markets and what its compensation is (risk premia).
- From the lectures on cross-sectional asset pricing, we have learned to improve the FM method to make it better handle the Errors In Variables (EIV) problem and use GMM to estimate it without EIV and robust to autocorrelation without assuming normality.
- However, we will also run into problems if we omit important control factors or if the risk factor is poorly measured or only weakly related to returns due to noise.
- As we know that there a hundreds of potential factors out there that might be relevant, this scenario doesn't seem unlikely.
- Luckily, recent advances in cross-sectional asset pricing and methods from the machine learning toolbox can help us mitigate these problems.
- Existing literature has typically ignored this bias.

4.2 Omitted variable bias

Assume true model, where z_i is the omitted variable:

$$Y = \beta X + \delta Z + E$$

x_i and z_i independent, driving the dependent variable y_i.
Running the regression using only X yields

$$\hat{\beta} = (X'X)^{-1} X'Y$$

giving us:

$$\mathbb{E}\left[\hat{\beta}|X\right] = \beta + \underbrace{(X'X)^{-1}\mathbb{E}\left[X'Z|X\right]\delta}_{\text{Bias}}$$

- The bias is a function of the sign of the covariance between x_i and z_i and the association between z_i and y_i.
- To capture the true estimate of β, we are strongly dependent on specifying the model correctly and accounting for all relevant variables.
  - This is where ML becomes useful.
- While for traded factors we can just take its sample mean as estimate of its risk premia, this is not possible for non-traded factors.

4.2.1 OVB in FM

Model with assumption of $\gamma_0 = 0$:

$$r_t^e = \beta_1 (\gamma_1 + v_{1t}) + \beta_2 (\gamma_2 + v_{2t}) + u_t$$

If we only use $g_t = v_{1t}$ as our subset of factors, we get OVB in both the first and second stage:

1. Estimate of β_1 is biased if the omitted factor is correlated with the included factor, with the magnitude depending on the size of this correlation.
2. Estimate of γ_1 is biased due to the former reason and that the cross-sectional regression omits $\hat{\beta}_2$ as regressor, with the magnitude of OVB depending on the size of the correlation among β_1 and β_2.

4.2.2 OVB in factor-mimicking portfolio

- OVB can also arise in risk premia estimate from the factor-mimicking portfolio if important base assets onto which g_t is projected are omitted.

Suppose we have base assets denoted $\breve{r}_t^e \subseteq r_t^e$ and regress $g_t = v_{1t}$ onto those assets with a constant. This yields ω^g and gives the mimicking portfolio

$$r_t^g = \omega^{g'} \breve{r}_t^e$$

Expected return for this portfolio:

$$\gamma_g^{MP} = \omega^{g'} \mathbb{E}[\breve{r}_t^e]$$

And since $\breve{r}_t^e \subseteq r_t^e$, this implies that:

$$\breve{r}_t^e = \breve{\beta} (\gamma + v_t) + \breve{u}_t$$

Using results from the OLS regressions, it turns out that

$$\gamma_g^{MP} = \left(\left(\breve{\beta} \Sigma^v \breve{\beta}' + \breve{\Sigma}^u \right)^{-1} \left(\breve{\beta} \Sigma^v e_1 \right) \right)' \breve{\beta} \gamma$$

For $\gamma_g^{MP} = \gamma_1$ to be true, we need two conditions:

1. $\breve{\Sigma}^u = 0$, which can be achieved if we e.g. have a well-diversified portfolio
2. $\breve{\beta}$ is invertible and $v_t = \breve{\beta}^{-1} \breve{r}_t^e$, such as if the true factors are fully spanned by the base assets.

If we omit some important base assets, though, we will violate either or both of these conditions, and does not get γ_1, causing a bias in the risk premia estimate.

4.3 Measurement error bias

- If the observable factor of interest g_t is measured with some error it will cause a bias in the estimated risk premium.
- This is especially common for non-traded factors which are often measured by some proxy.
- Hence we will get some bias to the risk premia estimator in both FM and factor-mimicking cases.

The econometrian observe g_t, where z_t is the measurement error orthogonal to the factors, but possibly correlated with u_t (the error term). This measurement will create attenuation bias in the estimated β in the time series regression.

$$g_t = v_{1t} + z_t$$

For Fama-MacBeth, this causes an attenuation bias in $\hat{\beta}_1$, which leads to a downward bias in $\hat{\gamma}_1$. For the factor-mimicking case, the relevant term expands to

$$\gamma_g^{MP} = \left(\left(\breve{\beta} \Sigma^v \breve{\beta}' + \breve{\Sigma}^u \right)^{-1} \left(\breve{\beta} \Sigma^v e_1 + \underbrace{\breve{\Sigma}^{z,u}}_{\text{New bias}} \right) \right)' \breve{\beta} \gamma$$

with $\breve{\Sigma}^{z,u} = Cov\left[z_t, u_t\right]$
If the measurement error is correlated with idiosyncratic risks of assets, we get a bias, even under the conditions above.

4.4 Three-pass methodology

- The three-pass methodology is a solution which can jointly tackle both the omitted variable and measurement error issues
- the objective is to estimate the risk premia of one or more factor g_t without having to observe all true factors v_t, as this is quite unrealistic.

$$\begin{aligned} r_t =& \beta\gamma + \beta v_t + u_t, \quad \mathrm{E}\left(v_t\right) = \mathrm{E}\left(u_t\right) = 0, \quad \text{and} \quad \mathrm{Cov}\left(u_t, v_t\right) = 0 \\ g_t =& \delta + \eta v_t + z_t, \quad \mathbb{E}\left[z_t\right] = 0, \mathrm{Cov}\left[z_t, v_t\right] = 0 \end{aligned}$$

The risk premium of g_t that we want to estimate is the expected excess return of a portoflio with $\beta = 1$ w.r.t. g_t and $\beta = 0$ w.r.t. all other factors, which implies that:

$$\gamma_g = \eta\gamma$$

Thus, in order to estimate the risk premia, we need some way to estimate the entire product $\eta\gamma$ or their individual constituents, where we are lucky enough to have contributions from Giglio and Xiu allowing for identification using a rotation invariance result.
Says that even if one just observes

$$\hat{v}_t = H v_t$$

with H any full rank p x p matrix, we can identify the product $\eta\gamma$ despite observing neither v_t nor H.
Define:

$$\begin{aligned} \hat{\eta} &= \eta H^{-1} \\ \hat{\gamma} &= H\gamma \\ \hat{\beta} &= \beta H^{-1} \end{aligned}$$

Now we can identify $\hat{\eta} = \eta H^{-1}$ as long as $\hat{v}_t$ is observed, since g_t and $\hat{v}_t$ are observed, so we can run a regression of g_t onto $\hat{v}_t$ including a constant, which estimate δ and $\hat{\eta}$ as the model reads

$$g_t = \delta + \hat{\eta}\hat{v}_t + z_t$$

Then we can also identify $\hat{\gamma} = H\gamma$ through a classical FM single cross-sectional regression using $\hat{\beta} = \beta H^{-1}$ as regressors and average r_t^e as dependent variables.
Now we can identify their product via

$$\hat{\eta}\hat{\gamma} = \eta H^{-1} H\gamma = \eta\gamma = \gamma_g$$

4.4.1 PCA

To actually identify these rotated factors, we can use Principal Components Analysis, which condenses a large data set into a smaller set of components that aims to capturing the most of the variation in data, giving us proxies for the unobserved factors v_t. Factors identified are not the true factors though, as any rotation of the factors would explain the same amount of variance, and thus we identify the factors up to a rotation, rather than actual factors.

4.4.2 Three-pass procedure

1. PCA step: Estimate rotated factors $\hat{v}_t$ via PCA on the full set of excess returns and obtain their factor loadings $\hat{\beta}$
2. Cross-sectional regression step: Run a standard cross-sectional OLS regression of average returns onto the estimated factor loadings from the prior step to obtain an estimate of the risk premia of the estimated latent factors $\hat{\gamma} = H\gamma$.

$$\hat{\gamma} = \left(\hat{\beta}'\hat{\beta}\right)^{-1}\hat{\beta}'\bar{r}$$

3. Time series regression step: Estimate $\hat{\eta} = \eta H^{-1}$ via a time-series regression of factor g_t onto $\hat{v}_t$
    - Run a time series regression of g_t onto the extracted factors from step 1 and then obtain the estimator, $\hat{\eta}$, and the fitted value of the observable factor, $\hat{G}$.

Then we can take the product $\hat{\eta}\hat{\gamma}$ of $\hat{\eta}$ and $\hat{\gamma}$ from steps two and three to get the risk premia of $g_t, \eta\gamma = \gamma_g$.

$$\hat{\eta}\hat{\gamma} = \eta H^{-1} H\gamma = \eta\gamma = \gamma_g$$

Essentially, the last step here is new compared to FM, and is critical in translating the uninterpretable risk premia of the latent factors to those that the economic theory predicts. It removes measurement error effects since we only use the effect coming through $\hat{\eta}$ and not the entire G. Thus we only extract the effect of the cleaned factor $\hat{G} = \hat{\eta}\hat{V} = \bar{G} - \hat{Z}$

- $\left(\overline{R}, \overline{V}, \overline{G}, \overline{U}, \overline{Z}\right)$ are demeaned matrices.
- $\hat{v}$: Rotated factors $\Rightarrow$ Principal component factors extracting out of all the latent underlying true factors that explain variation in our sample.
- $\hat{\beta}$: Loadings for the principal component factors that we extract $\hat{v}$.
- $\hat{\gamma}$: Risk premia of estimated latent factors.
- $\hat{\eta}$: Explains the relationship between $\hat{v}_t$ and g_t.
- $\hat{G}$: Is the cleaned version of the observable factor G which means that measurement error is removed.
- One important property is that the estimation for each observable factor is done seperate $\Rightarrow$ g_t can be multidimensional.

4.5 Interpretation

- We can then interpret our three-pass estimator as a factor-augmented cross-sectional regression estimator.
- Our three-pass procedure can also be interpreted as a mimicking-portfolio estimator, in which the PCs themselves are the portfolios on which gt is projected. This is an ideal choice of portfolios that ensures that the estimator is consistent.

4.5.1 Methods support

Extension of both FM and factor mimicking, actually presenting a convergence of the two, resulting in them giving the same results, and lends support to both methods, for FM by ensuring no OVB in both first and second stage and for factor-mimicking by exactly averaging out noise and satisfying the conditions for no OVB.

4.5.2 Asymptotic theory

- The main point is that the three-pass estimator is asymptotically consistent as the number of test assets (N) and time-periods (T) goes towards infinity when including the right number of latent factors p (length of v).

Consistency: If we have that $\hat{p} \xrightarrow{p} p$, then as $N, T \to \infty$, we have that:

$$\hat{\eta} \xrightarrow{p} \eta H^{-1}$$
$$\hat{\gamma} \xrightarrow{p} H\gamma$$
$$\hat{\eta}\hat{\gamma} = \hat{\gamma}_g \xrightarrow{p} \gamma_g$$

implying that we get the right estimates as long as we consistently estimate the correct number of latent factors.

Asymptotic normality:
If $\hat{p} \xrightarrow{p} p$, then, as $N, T \to \infty$ together with $T^{1/2}N^{-1} \to 0$, we get that

$$T^{1/2}\hat{\gamma}_g \xrightarrow{d} N(\gamma_g, \Phi)$$

4.6 Ridge regression

Alternative to PCA, but sucks compared to PCA, as ridge puts weight on all assets and will reflect some of the noise inherent in the assets, whereas PCA averages out the noise by forming the optimal linear combinations that extract the relevant signal.

4.7 Testing strength

Weak factors are observable factors that are only weakly reflected in the cross-section of test assets, posing challenges for econometric techniques. In the model:

$$g_t = \delta + \eta v_t + z_t$$

an $\eta \approx 0$ indicates that either measurement error dominates, or the factor is not strong. Then the loading would be weak, meaning it is poorly estimated with little cross-sectional variation.
We can test $\mathbb{H}_0 : \eta = 0$ vs. $\mathbb{H}_1 : \eta \neq 0$ by formulating a conventional Wald test as per

$$\hat{W} = T\hat{\eta}\left(\left(\hat{\Sigma}^{v}\right)^{-1}\hat{\Pi}_{11}\left(\hat{\Sigma}^{v}\right)^{-1}\right)^{-1}\hat{\eta}' \xrightarrow{d} \chi^2_{\hat{p}}$$

Using R^2 from the same regression as a measure of strength

$$\hat{R}^2_g = \frac{\hat{\eta}\hat{V}\hat{V}'\hat{\eta}'}{\overline{GG}'} \xrightarrow{p} R^2_g$$

where $R^2_g$is the true version.

A factor is said to be strong with strength $\hat{R}^2_g$ if we reject $\mathbb{H}_0 : \eta = 0$, meaning that $\hat{W}$ exceeds the $\chi^2_{\hat{p}}(1-\alpha)$ percentile for α significance level.

- Giglio and Xiu (2021) finds:
  - A rejection of the null indicates that g_t is a strong factor for the cross section of test portfolios.

5 Giglio and Xiu (2021)

Main points:

- The main advantage of the three-pass methodology is that it provides a systematic way to tackle the concern that the model predicted by theory is misspecified because of omitted factors.
- Rather than relying on arbitrarily cho- sen control factors or computing risk premia only on subsets of the test assets, our methodology utilizes the large dimension of testing assets avail- able to span the factor space.
- It also explicitly takes into account the possibility of measurement error in any observed factor.

5.1 Taming the factor zoo

- Examine how machine learning algorithms can help us to determine whether a new factor is truly new or redundant.
- Do a new factor contain a marginal contribution when explaining the cross-section of returns.
- Example why this is relevant:
  - Heston and Sadka (2008) introduces seasonality factor $\Rightarrow$ highly significant in FF3 model.
  - When including momentum factor, this seasonality factor seems though to be redundant.
- Using a single LASSO, ror any finite sample, we cannot ensure that the method selects the true model leading to a potential OVB
- They propose a regularized two-pass cross-sectional regression approach to establish the asset pricing contribution of a factor g_t relative to a set of control factors h_t, where the potential control set can have high dimensionality and include useless or redundant factors.
- Their two-step method is able to produce correct inference by overcoming the model selection mistakes that necessarily arise when applying statistical selection methods.
- Several newly proposed factors (especially different versions of profitability) are useful in explaining asset prices, even after accounting for the large set of factors proposed prior to 2012.
- They show that applying our test recursively over time would have deemed only a small number of factors proposed in the literature significant.
  - Their results differ starkly from the conclusions one would obtain simply by using the risk premia of the factors or the standard Fama-French three-factor model as a control

To guard against OVB Feng et al. (2020) adopt a double-selection step prior to standard FM analysis which should capture the missed factors from the LASSO.

Double selection:
First selection: Search for factors in h_t (set of observable factor candidate controls) that are useful for explaining the cross-sectional variation in expected returns. Can be done using classical LASSO.
Second selection: Search for factors in h_t that are useful for explaining the cross-sectional variation in exposures to the risk factor g_t (covariances $\text{Cov}\left[g_t, r_{it}^e\right]$) - also by LASSO.
Then the selected variable $\tilde{h}_t \subseteq h_t$ is chosen so as to minimize OVB, and finalized by including $\tilde{h}_t$ as controls in a classical FM analysis.

Fama-Macbeth:
Standard second-stage single cross-sectional FM regression using estimated covariances between r_{it}^e and both g_t and $\tilde{h}_t$ instead of β_is. If β_is had been used in step 1 and 2 we examine whether a new risk factor generates an adjusted return that cannot be explained by existing factors.

Happens to require data on all relevant factors which is very troublesome compared to just having a large set of portfolios as in the three-pass approach of Giglio and Xiu (2021)
Further examining the robustness of the double selection model Feng et al. (2020) finds that the selected variables depends on the seeding of the random number generator used for the k-fold cross-validation used for training the LASSO.

- LASSO:
  - Statistical method that select a set of covariates for a given model. Was introduced to improve the prediction accuracy and interpretability of regression models.