2. Cross-sectional asset pricing

Contents

1 Learning objective:

- **Reading list:** Campbell (2017, ch. 4), Goyal (2012), Lecture notes
- The stochastic discount factor (SDF), its existence and use in asset pricing, its implications for arbitrage and market completeness, and its relation to the investor's marginal utility
- Econometric methods and techniques for estimating and testing risk premia in the cross- section of asset returns
- Discuss and estimate the SDF using common empirical methods, evaluate its ability to price the cross-section and conduct valid inference, and reflect on the findings and their implications

2 Suggested presentation 10 min

1. Empirical

Part I
Short version

3 Cross-sectional asset pricing

3.1 Introduction

- Cross-sectional rergessions and time series regression are primary approach for testing asset pricing models
- Can be represented with β or SDF, with one implying the other
- Consider two methods:
  - Very popular and easily implementable Fama-MacBeth method
  - GMM method that accounts for some of the errors with FMB

3.2 Fama-MacBeth

3.2.1 First-stage

$$R_{it} - R_{ft} = \alpha_i + \beta_i f_t + \varepsilon_{it}$$

Risk factors chosen for multiple reasons:

- Economical intuition
- Statistical findings
- Firm characteristics

Get $\hat{\beta}$ estimates.
Can be done full-sample as above $\Rightarrow \hat{\beta}_i$, or on a rolling basis $\Rightarrow \hat{\beta}_{it}$

3.2.2 Second-stage

$$R_{it} - R_{ft} = \gamma_0 + \gamma_{it}\hat{\beta}_i + \eta_{it}$$

The risk factor is then obtained as the average of the j estimates.

$$\hat{\gamma}_j = \frac{1}{T}\sum_{t=1}^{T}\hat{\gamma}_{jt}$$

Get $\hat{\gamma}$ estimates by regressing on estimated β values.
Can be done in single cross-sectional regression to get same coefficients, but this gives very wrong OLS standard errors.
$\overline{R_{it} - R_{ft}} = \gamma_0 + \gamma\hat{\beta}_i + \eta_i$

3.3 EIV

- Huge drawback with FM
  - Caused by estimations from unobservable factors, which is generally the case for non-traded factors.
  - Estimations used in second-stage $\Rightarrow$ overstated precision
  - Mitigations:
    * Group into portfolios to average out noise
    * Use larger time-series

3.3.1 Shanken correction

- Scale the OLS errors upwards to reflect overstated precision of $\hat{\gamma}$
- Different representations depending on the type of FM used
  - 1. Full-sample, constant βs, t-by-t

$$Var_{EIV}\left[\hat{\gamma}\right] = T^{-1}\left(\left(1+c\right)\left(TVar\left[\hat{\gamma}\right] - \tilde{Var}\left[f_t\right]\right) + \tilde{Var}\left[f_t\right]\right)$$

    2. Full-sample, constant βs from first stage, single cross-sectional stage

$$Var_{EIV}\left[\hat{\gamma}\right] = T^{-1}\left(\left(1+c\right)TVar\left[\hat{\gamma}\right] + \tilde{Var}\left[f_t\right]\right)$$

    3. Rolling βs, estimated over y years, first stage t-by-t

$$Var_{EIV}\left[\hat{\gamma}\right] = T^{-1}\left(\left(1+c^*\right)TVar\left[\hat{\gamma}\right] + \tilde{Var}\left[f_t\right]\right)$$

    Where:

$$\begin{aligned} c &= \hat{\gamma}'\tilde{Var}\left[f_t\right]^{-1}\hat{\gamma} \\ c^* &= c\left(1 - \frac{\left(y-1\right)\left(y+1\right)}{3yT/m}\right) \end{aligned}$$

3.4 GMM

- Even with Shanken corrections, FM doesn't account for autocorrelation, & tend to assume normality of regression erros, which can wrong
- Solution $\Rightarrow$ estimate jointly with GMM
- Define two sets of moments, corresponding to the first and second stage of FM, but calculated jointly
  - Joint estimate "internalizes" any EIV $\Rightarrow$ long-run covariance matrix S will caputre directly the effect of generated regressors.
  - Can be amended nonparametrically to account for autocorrelation and heteroskedasticity
- Moment conditions:
-

$$\mathbb{E}\begin{bmatrix} \underbrace{R_t^e - \alpha - \beta f_t}_{\varepsilon_t} \\ \underbrace{\left(R_t^e - \alpha - \beta f_t\right)\otimes f_t}_{\varepsilon_t \otimes f_t} \\ \underbrace{R_t^e - \gamma_0 - \gamma\beta}_{\eta_i} \end{bmatrix} = \begin{bmatrix} 0_{Nx1} \\ 0_{NKx1} \\ 0_{Nx1} \end{bmatrix}$$

- Amend moments to fix overidentification by pre-multiplying with:
  - $e = \begin{bmatrix} I_{N(K+1)} & 0_{N(K+1)xN} \\ 0_{(K+1)xN(K+1)} & \chi' \end{bmatrix}$

        * $\chi' = (\iota_{Nx1}, \beta)$
    - $eg = 0$
        * Reproduces two-pass FMM

- Standard errors obtained from this are now robust to EIV, autocorrelation, and heteroskedasticity!
    - $\theta' = \left(\alpha', vec\left(\beta\right)', \gamma_0, \gamma\right)'$

3.5 Further critique

- These simple methods are very easy to apply, and finding high cross-section R^2 with small pricing errors with little economic meaning is senseless and leads to unnecessary factor zoo problems.
- Lewellen's prescriptions:
    1. Use more than FF size-value portfolios, and add SMB and HML to see whether they drive out new factors
    2. Seriously consider sign, magnitude and significance of cross-sectional coefficients
        (a) Test $\gamma_0 = 0$ if implied
    3. Report confidence intervals
- Other good ideas:
    - Bootstrap/placebo distribution of R^2
    - Sufficient amount of robustness checks with other sample periods or test assets

Part II
Long version

4 Introduction

5 Cross-sectional Asset Pricing: Theory and Estimation

5.1 Introduction

The objective of this subject is that we want to be able to test whether any given factor is priced in a cross-section of assets.

Time-series regressions and cross-sectional regressions make up the primary approach for testing asset pricing models, which typically takes offset in either a β representation or a linear factor model in SDF representation, with the two representations being intimately linked in that one can lead to the other and vice versa. There are generally two methods we will consider for doing this: The Fama-MacBeth procedure and the Generalized Method of Moments (GMM).

5.2 Fama-MacBeth

Cross-sectional regression method consisting of two steps. We work with it in the β representation.

5.2.1 First-stage regression

Here we estimate the β_i coefficients for risk factors using a full sample time-series OLS regression:

$$R_{it} - R_{ft} = \alpha_i + \beta_i f_t + \varepsilon_{it}$$

The risk factors involved in such a regression can be chosen for a variety of reasons, commonly based on economic intuition, statistical extraction or firm characteristics.
Having run this regression, we obtain $\hat{\beta}_i$. One can either use a full-sample estimation for this which assumes a constant factor loading, or we can extend it to incorporate rolling βs.

Rolling βs: If one is interested in allowing for time-variation in the estimated β's, then these can be estimated using a rolling window.
Done by simply running a first-stage regression at every time point t for $t = M, ..., T$, where M is the length of the rolling window.
This generates $\hat{\beta}_{it}$, which can simply be included in FM regressions by using $\hat{\beta}_{it}$ for the $t+1$'th cross-sectional regression instead of always using $\hat{\beta}_i$.

5.2.2 Second-stage regression(s)

Run the cross-sectional regression of all assets against the estimated betas from the first step in order to obtain an estimate of the risk premia, as:

$$R_{it} - R_{ft} = \gamma_{0t} + \gamma_t \hat{\beta}_i + \eta_{it}$$

Here, there are T cross-sectional regressions to run, providing us with time series estimates of $\{\hat{\gamma}_{0t}, \hat{\gamma}_t\}$, which can be used to form estimates of the risk factors $\hat{\gamma}_0$ and $\hat{\gamma}$ as:

$$\hat{\gamma}_j = \frac{1}{T} \sum_{t=1}^{T} \hat{\gamma}_{jt}$$

Can also be run as a single cross sectional regression based on rational expectations of investors which provides the same γ estimates as the t-by-t procedure. Note that the OLS standard errors are <u>very wrong</u>.

$$\overline{R_{it} - R_{ft}} = \gamma_0 + \gamma \hat{\beta}_i + \eta_i$$

$$\overline{R_{it} - R_{ft}} = T^{-1} \sum_{t=1}^{T} R_{it} - R_{ft}$$

5.2.3 Variance and hypothesis testing

Variance of each risk premium estimate:

$$\text{Var}\left[\hat{\gamma}_j\right] = \frac{1}{T^2}\sum_{t=1}^{T}\left(\hat{\gamma}_{jt} - \hat{\gamma}_j\right)^2$$

Entire covariance matrix:

$$\text{Var}\left[\hat{\gamma}\right] = \frac{1}{T^2}\sum_{t=1}^{T}\left(\hat{\gamma}_t - \hat{\gamma}\right)\left(\hat{\gamma}_t - \hat{\gamma}\right)'$$

What we want to test with this is the hypothesis that

$$\mathbb{H}_0 : \gamma_j = 0$$

Which can be tested with the conventional t-statistic:

$$t\left(\gamma_j\right) = \frac{\hat{\gamma}_j}{\sqrt{Var\left[\hat{\gamma}_j\right]}} \xrightarrow{d} N\left(0,1\right), \qquad \text{as } T \to \infty$$

5.3 Drawbacks

There are several drawbacks associated with the Fama Macbeth approach:

- The betas used in the cross-sectional regression are estimates which contain estimation errors, causing an overstated precision of the risk premia estimates.
  - This gives rise to the so-called "errors in variables" problem.
  - his can be mitigated by grouping assets into portfolios to average out noise, or by using larger time-series, but this will hardly be satisfactory, why we need to "correct" the errors.
  - We can overcome this challenge using the Schanken correction.
- The Fama Macbeth does not account for autocorrelation and heteroskedasticity.
- We can turn to GMM which fixes the EIV as well as the issues with autocorrelation and heteroskedasticity.

5.3.1 Shanken correction

A solution is the Shanken correction, where we scale the OLS standard errors upwards to reflect the overstated precision of $\hat{\gamma}$.

The correction itself depends on which type of FM approach we use. These effects can severely impact conclusions. It should be noted that the need for corrections depend on the factors used. Firm characteristics are for example typically directly observed and don't need to be estimated. Can be fixed by augmenting the $\tilde{Var}\left[f_t\right]$ as to include zero rows and colums at the place of the characteristics.
Shanken-corrected covariance matrices:
1: Full-sample, constant βs from first stage in a t-by-t cross-sectional stage:

$$\text{Var}_{EIV}\left[\hat{\gamma}\right] = T^{-1}\left(\left(1+c\right)\left(T\cdot\text{Var}\left[\hat{\gamma}\right] - \widetilde{\text{Var}}\left[f_t\right]\right) + \widetilde{\text{Var}}\left[f_t\right]\right)$$

2: Full-sample (constant) βs from first stage in a single cross-sectional stage

$$\text{Var}_{EIV}\left[\hat{\gamma}\right] = T^{-1}\left(\left(1+c\right)T\cdot\text{Var}\left[\hat{\gamma}\right] + \widetilde{\text{Var}}\left[f_t\right]\right)$$

3: Rolling βs, estimated over y years with m data points per year, from first stage in a t-by-t cross-sectional stage.

$$\text{Var}_{EIV}\left[\hat{\gamma}\right] = T^{-1}\left(\left(1+c^*\right)T\cdot\text{Var}\left[\hat{\gamma}\right] + \widetilde{\text{Var}}\left[f_t\right]\right)$$

Where:

$$c = \hat{\gamma}' \widetilde{\mathrm{Var}}\left[f_t\right]^{-1} \hat{\gamma}$$
$$c^* = c\left(1 - \frac{(y-1)(y+1)}{3yT/m}\right)$$

$\mathrm{Var}\left[f_t\right]$ is the sample covariance matrix of the risk factors. $\widetilde{\mathrm{Var}}\left[f_t\right]$ is the *(expanded)* covariance matrix with zeroes in the top row and first column corresponding to the intercept.

5.4 GMM approach

It is possible to combine TSR and CSR into a unified generalized method of moments (GMM) framework.

- Even with Shanken corrections, the FM method still doesn't account for the presence of autocorrelation.
- Tend to assume normality of regression errors, which can be wrong.
- Using GMM estimation will overcome (almost) all issues such as:
  - Errors in variables
  - Autocorrelation
  - Heteroskedasticity
  - More mild assumptions

The main idea is to define two sets of moments:

1. First set matches the time series stage for obtaining β
2. Second set matches the cross-sectional stage for obtaining γ

- Merging these two sets of moments "internalizes" any EIV due to β and γ being estimated jointly
- the long-run covariance matrix of moments, S will capture directly the effect of generated regressors
- and can be amended non-parametrically to account for autocorrelation and heteroskedasticity through HAC.

This captures almost all issues in one set of moments, is essentially still a FM regression analysis, but with an additinoal layer added to provide proper and accurate standard errors accounting for both EIV, autocorrelation, and heteroskedasticity, while being much more milder in assumptions.

GMM Technicals
Time-series regression in vector form:

$$R_t^e = \alpha + \beta f_t + \varepsilon_t$$

OLS conditions *(for a univariate dependent variable y, regressors x, and coefficients θ)*: The error term is mean zero and that it is uncorrelated with the regressors. $(y - \theta x)$ denotes the residual.

$$\mathbb{E}\left[y - \theta x\right] = 0$$
$$\mathbb{E}\left[(y - \theta x)\, x\right] = 0$$

The OLS estimation of the time series stage then maps into the following set of moments (N number of assets, T periods and K factors):

$$\mathbb{E}\begin{bmatrix} R_t^e - \alpha - \beta f_t \\ (R_t^e - \alpha - \beta f_t) \otimes f_t \end{bmatrix} = \mathbb{E}\begin{bmatrix} \varepsilon_t \\ \varepsilon_t \otimes f_t \end{bmatrix} = \begin{bmatrix} 0_{Nx1} \\ 0_{NKx1} \end{bmatrix}$$

equalling $N + NK$ moment conditions. The sum is exactly identified since we N many α's and NK many β's to estimate - reducing to the analytic solution of the OLS for each i, while the Kronecker product ensures that all error terms are uncorrelated with all risk factors.
With parameters exactly identified, the analytical solution is OLS and we might as well just run simple OLS regressions equal to the first stage of FM, getting the same parameter estimates.

From the β representation of the K-factor asset pricing model describes

$$\mathbb{E}\left[R_t^e\right] = \gamma_0 + \gamma\beta$$

In many cases there will be no constant γ_0 and the restriction $\gamma_0 = 0$ could be imposed, but will often be included as a testable restriction. From this, we can define the second set of moments, with the following moment conditions:

$$\mathbb{E}\left[R_t^e - \gamma_0 - \gamma\beta\right] = 0_{Nx1}$$

With joint moment conditions given by

$$\mathbb{E}\begin{bmatrix} R_t^e - \alpha - \beta f_t \\ (R_t^e - \alpha - \beta f_t) \otimes f_t \\ R_t^e - \gamma_0 - \gamma\beta \end{bmatrix} = \mathbb{E}\begin{bmatrix} \varepsilon_t \\ \varepsilon_t \otimes f_t \\ \eta_i \end{bmatrix} = \begin{bmatrix} 0_{Nx1} \\ 0_{NKx1} \\ 0_{Nx1} \end{bmatrix}$$

This system is overidentified as N moments are added and we only need to estimate $K+1$ additional parameters.

This system will generally **not** reproduce OLS, and thus the moments need to be amended. Do that by defining the following matrix:

$$e = \begin{bmatrix} I_{N(K+1)} & 0_{N(K+1)xN} \\ 0_{(K+1)xN(K+1)} & \chi' \end{bmatrix}$$

with $\chi = (\iota_{Nx1}, \beta)$ and ι a N-vector of ones
When pre-multiplying e onto the joint moment conditions, we maintain the first $N + NK$ conditions but weight each of the last N moment conditions by χ.
We then get total OLS type moment conditions:

$$\begin{bmatrix} I_{N(K+1)} & 0_{N(K+1)xN} \\ 0_{(K+1)xN(K+1)} & \chi' \end{bmatrix} \mathbb{E}\begin{bmatrix} R_t^e - \alpha - \beta f_t \\ (R_t^e - \alpha - \beta f_t) \otimes f_t \\ R_t^e - \gamma_0 - \gamma\beta \end{bmatrix} = \begin{bmatrix} 0_{Nx1} \\ 0_{NKx1} \\ 0_{Nx1} \end{bmatrix}$$

equivalent to $eg = 0_{(N+1)(K+1)}$, and this reproduces the two-pass FM estimates.

Using this, we can define the GMM recipe for asset pricing quite simply as:

1. Estimate β_i for all i via standard FM time series stage regressions
2. Estimate γ_j for all j in K via standard FM single cross-sectional regression
3. Use the definition of e above and the joint moment condition system to obtain standard errors that are robust to EIV, autocorrelation, and heteroskedasticity.

To compute such standard errors, define:

$$\theta' = \left(\alpha', vec\left(\beta\right)', \gamma_0, \gamma\right)'$$

which contains all $(N+1)()$
We are interested in the **covariance of** $\hat{\theta}$

$$Var\left[\hat{\theta}\right] = T^{-1}\left(eD\right)^{-1} eSe'\left(eD\right)^{-1'}$$

where $D = \mathbb{E}\left[\partial g\left(\theta\right)/\partial\theta\right]$ equal to:

$$D = -\begin{bmatrix} \begin{bmatrix} 1 & \mathbb{E}\left[f_t'\right] \\ \mathbb{E}\left[f_t\right] & \mathbb{E}\left[f_t f_t'\right] \end{bmatrix} \otimes I_N & 0_{N(K+1)x(K+1)} \\ \begin{bmatrix} 0 & \gamma' \end{bmatrix} \otimes I_N & \chi \end{bmatrix}$$

where $S$is the long-run covariance matrix equal to

$$S = \sum_{s=-\infty}^{\infty}\left[\begin{bmatrix} R_t^e - \alpha - \beta f_t \\ (R_t^e - \alpha - \beta f_t) \otimes f_t \\ R_t^e - \gamma_0 - \gamma\beta \end{bmatrix}\begin{bmatrix} R_{t-s}^e - \alpha - \beta f_{t-s} \\ \left(R_{t-s}^e - \alpha - \beta f_{t-s}\right) \otimes f_{t-s} \\ R_{t-s}^e - \gamma_0 - \gamma\beta \end{bmatrix}'\right]$$

5.5 Traded vs. non-traded factors and estimating SDF loadings

5.5.1 Traded vs. non-traded

- Some factors are returns themselves such market risk premium, SMB and HML.
  - We don't need to estimate their risk premium in a cross-sectional stage, but can suffice with taking the sample mean.
  - Lewellen (2010) emphasize that one diagnostic test of an asset pricing model with traded factors is that the risk premia estimate from CSR and TSR (or sample mean) should be statistically indistinguishable.
- With non-traded factors, we need to estimate risk premia using the cross-sectional stage.

5.5.2 Estimating SDF loadings

- We have worked with the β representation so far.
- We could also have estimated the parameters of the SDF (called SDF loadings) denoted by b.
- Most asset pricing models can be recast in stochastic discount factor (SDF) frame- work (Cochrane 2005).
- **SDF Generally**
  - The SDF can be seen as the factor which equalizes the discounted future cash flow and the price of an asset.
  - A positive SDF exists if and only if markets are free of arbitrage.
  - Given no-arbitrage, the SDF is unique if and only if markets are complete
  - If an SDF exists, we can always find a β representation for assetreturns, and vice versa
  - A SDF, M_t, is a random variable with the following properties:
    * M_t has a finite variance.
    * M_t is strictly positive.
    * The price of asset i is given by:

$$p_{i,t}\left(X_{i,t+1}\right) = \mathbb{E}_t\left(M_{t+1}X_{i,t+1}\right)$$

We assume an SDF of the form

$$M_t = 1 - b' f_t$$

and can then form standard GM moment conditions form the fundamental equation of asset pricing as

$$\mathbb{E}\left[\left(1 - b' f_t\right) R_t^e\right] = 0_{N\times 1}$$

Having

$$D = \mathbb{E}\left[R_t^e f_t'\right]$$

which can be estimated simply as

$$T^{-1}\sum_{t=1}^{T} R_t^e f_t'$$

noting that

$$S = \sum_{s=-\infty}^{\infty} \mathbb{E}\left[\left(R_t^e - b' f_t R_t^e\right)\left(R_{t-s}^e - b' f_{t-s} R_{t-s}^e\right)'\right]$$

Then we can get the following two SDF loadings depending on which matrix we are using for the weighting matrix, with identity and optimal S^{-1} yield respectively:

$$\hat{b}_1 = \left(D'D\right)^{-1} D'\overline{R^e}$$
$$\hat{b}_2 = \left(D'S^{-1}D\right)^{-1} D'S^{-1}\overline{R^e}$$

with

$$
\begin{aligned}
Var\left[\hat{b}_1\right] &= T^{-1}\left(D'D\right)^{-1}D'SD\left(D'D\right)^{-1} \\
Var\left[\hat{b}_2\right] &= T^{-1}\left(D'SD\right)^{-1}
\end{aligned}
$$

- **The stochastic discount factor (SDF), its existence and use in asset pricing, its implications for arbitrage and market completeness, and its relation to the investor's marginal utility**
  - The SDF can be seen as the factor which equalizes the discounted future cash flow and the price of an asset.
  $$p_{i,t}\left(X_{i,t+1}\right) = \mathbb{E}_t\left(M_{t+1}X_{i,t+1}\right)$$
  - Ruling out arbitrage opportunities guarantees the existence of a positive SDF.
  - Near-arbitrages opportunities imply extreme volatile SDF for the investors $\Rightarrow$ extreme volatile marginal utility for rational investors.
  - If markets are incomplete? Then the SDF is not unique, but under some very weak condtions, there exists a positive SDF that can price all financial assets.
  - The positivity of the SDF ensures positive marginal utility of consumption and a positive price of assets providing a positive payoff
  - Given no-arbitrage, the SDF is unique if and only if markets are complete.
  - A postive SDF exists if and only if prices admits no arbitrage.
- Econometric methods and techniques for estimating and testing risk premia in the cross- section of asset returns
- **Discuss and estimate the SDF using common empirical methods, evaluate its ability to price the cross-section and conduct valid inference, and reflect on the findings and their implications**
  - Imposing a utility function we can derive the SDF in the consumption based framework
  - The return on any asset is:
    * The risk-free return
    * A term that informs about the co-variation between the SDF and returns
  - Within the consumption based asset-pricing framework, we can write the SDF as:
  $$M_{t+1} = \delta\frac{u'\left(c_{t+1}\right)}{u'\left(c_t\right)}$$

5.5.3 Critique

Lewellen et al. (2010) finds:

- Too easy to find a high cross-sectional R^2 with small pricing errors, often with little economic meaning - which cannot provide much support for a proposed model by itself.
- More than a sampling error, though sampling errors exarcebate the problem.

Three of Lewellen's six prescriptions for improving asset pricing analyses are:

1. Expand the test of set portfolios to encompass more than FF size-value portoflios. Relatedly, one can add SMB and HML to see whether they drive out the new factors.
2. Take the magnitude, sign and significance of the cross-sectional coefficients seriously
   (a) If $\gamma_0 = 0$ is implied, be sure to test this and comment!
   (b) If a model implies that γ is equal to average factor excess returns for traded factors, make sure to evaluate, otherwise indicating model misspecification
3. Report confidence intervals for the cross-sectional R^2. Will usually provide wide bands, as it is an estimated metric and should be considered as such.

Other good ideas:

1. Test for model misspecification via adequate cross-sectional dispersion in estimated βs
2. Bootstrap or placebo distribution of R^2
3. Sufficient amount of robustness checks such as with other sample periods or test assets.
4. Most importantly, always have a strong economic motivation for why the model works.