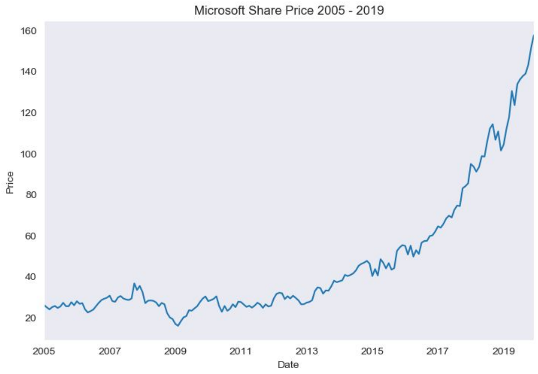
Estimating liquidity betas of stocks using a multiple linear regression model

General Assembly - Data Science - Part Time

Ravi Verma

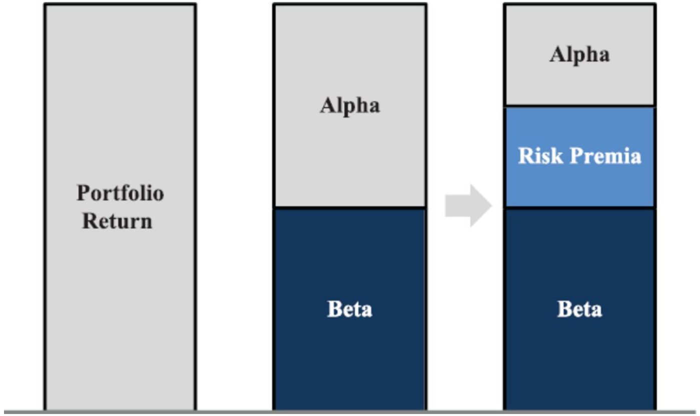
Introduction - shares

- Many companies around the world allow you to buy a share in them
- These ‘shares’ represent a tiny percentage of the company and are normally listed on a exchange
- They normally give you rights such as voting at company meetings on particular topics or receiving benefits such as dividends
- The shares of these companies have a price that move and up and down over time
- Just like other products, you can make a profit/return from buying and selling at the right time



Introduction - factors

- Returns of shares can often be visualised as a set of building blocks
- These building blocks are known as alpha and beta
- Beta can be thought of the return from the market e.g. the entire universe of shares
- Alpha can be thought of the return idiosyncratic to that particular share
- However over many years of quantitative research there are other ‘factors’ believed to significantly contribute to share returns, also known as ‘risk premia’



Introduction - regression

- Regression models can estimate how influential these factors are on stock returns
- We can think of stock price as the dependent variable
- The independent variables would be the factors
- These factors have been studied by many academics over a long time. In particular, Fama and French produced a seminal model, the three-factor model which describes market, size and value as significant factors that help explain stock returns
- Liquidity was shown to be a significant factor by the academics Pastor and Stambaugh
- Putting this all together we get the following multiple linear regression equation:

$$r_{i,t} = \beta_i^0 + \beta_i^{\mathcal{L}} \mathcal{L}_t + \beta_i^{\mathrm{M}} \mathrm{MKT}_t + \beta_i^{\mathrm{S}} \mathrm{SMB}_t + \beta_i^{\mathrm{H}} \mathrm{HML}_t + \epsilon_{i,t},$$

Introduction - my project

- I am going to use the Fama-French and Pastor-Stambaugh factor returns for market, value, size and liquidity, respectively. These are available from their respective academic websites (free to use)
- I have selected 89 stocks from the S&P 100 Index. I am using monthly returns from 2005 - 2019. This is 15 years worth of data or 180 months per stock. I have used FactSet to download these returns, but many if not all would be available from Yahoo Finance (free to use)
- The factor I am most interested in is liquidity and therefore my results will focus exclusively on that factor
- To estimate the relationship between factors and stock returns I will use OLS regression, partial regression and rolling regression techniques

Introduction - data sources

https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

http://finance.wharton.upenn.edu/~stambaug/

mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

RAPHY
CULUM VITAE
NG PAPERS
LIBRARY
RESEARCH RETURNS

Current Research Returns

In August 2019, we added emerging markets portfolios to the bottom of the page. The global portfolios and factors have been renamed to developed.

	March 2020	Last 3 Months	Last 12 Months
Fama/French 3 Research Factors			
Rm-Rf	-13.39	-20.57	-10.60
SMB	-5.13	-6.16	-11.68
HML	-14.11	-20.13	-24.26

Robert F. Stambaugh

- entry in Wharton Guide to Faculty
- curriculum vitae

Portfolio Optimization:

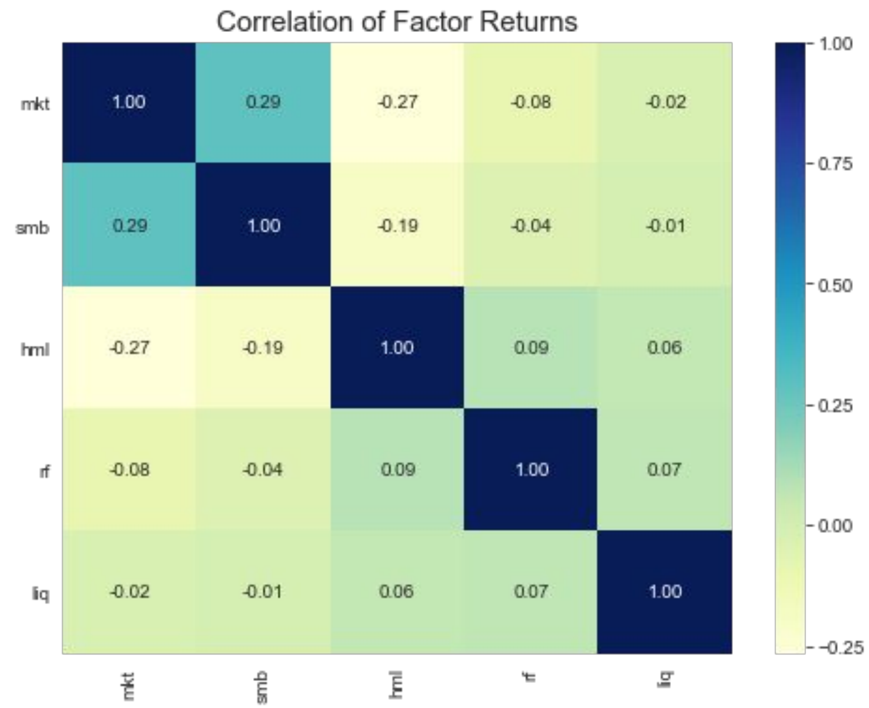
- program and description
- some filled-in input forms for the optimization program, used as examp

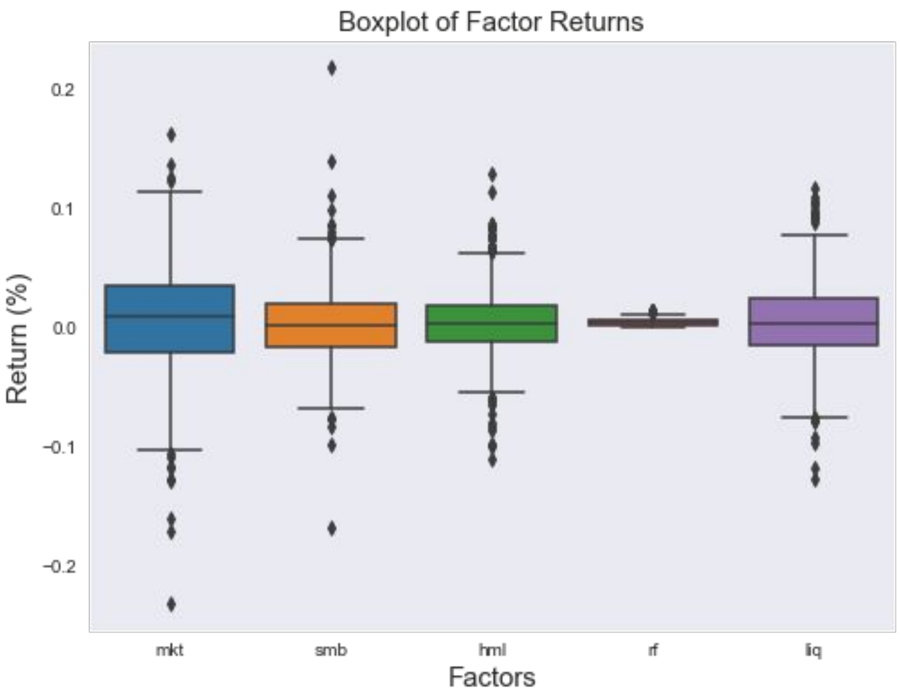
Data:

- Pastor-Stambaugh liquidity series
- Plot of the aggregate liquidity level.

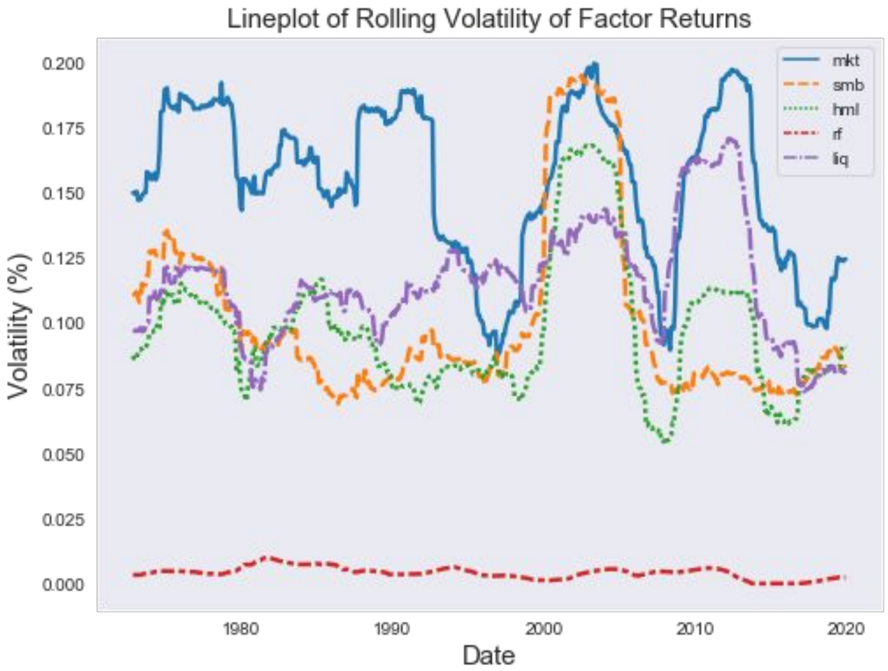
- Stambaugh and Yuan "Mispricing Factors" data (1/1963 - 12/2016):
  - monthly factors
  - daily factors
  - 11 anomaly returns

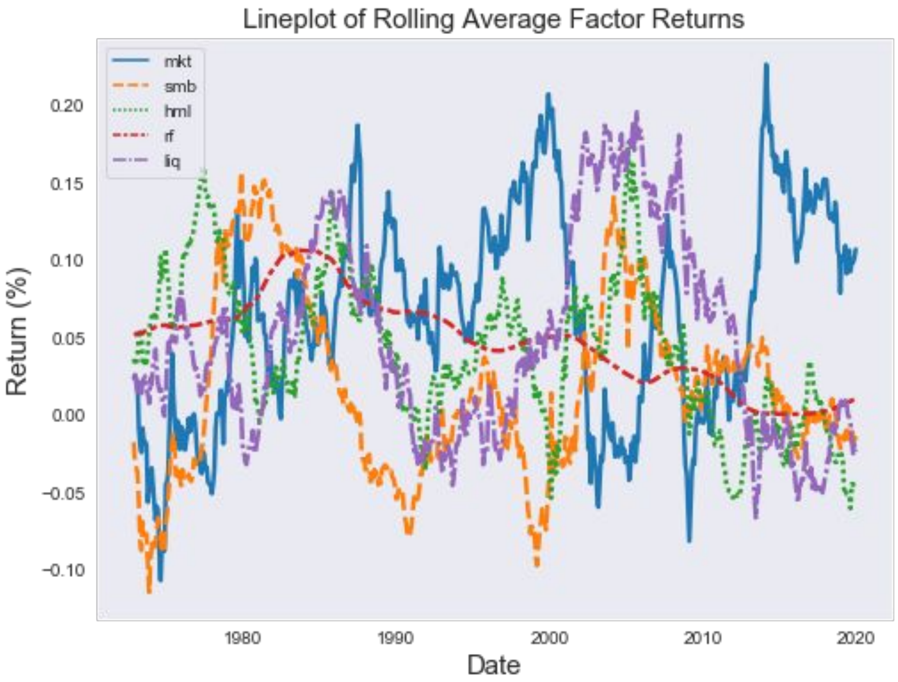
Exploratory Data Analysis - factors (part 1)



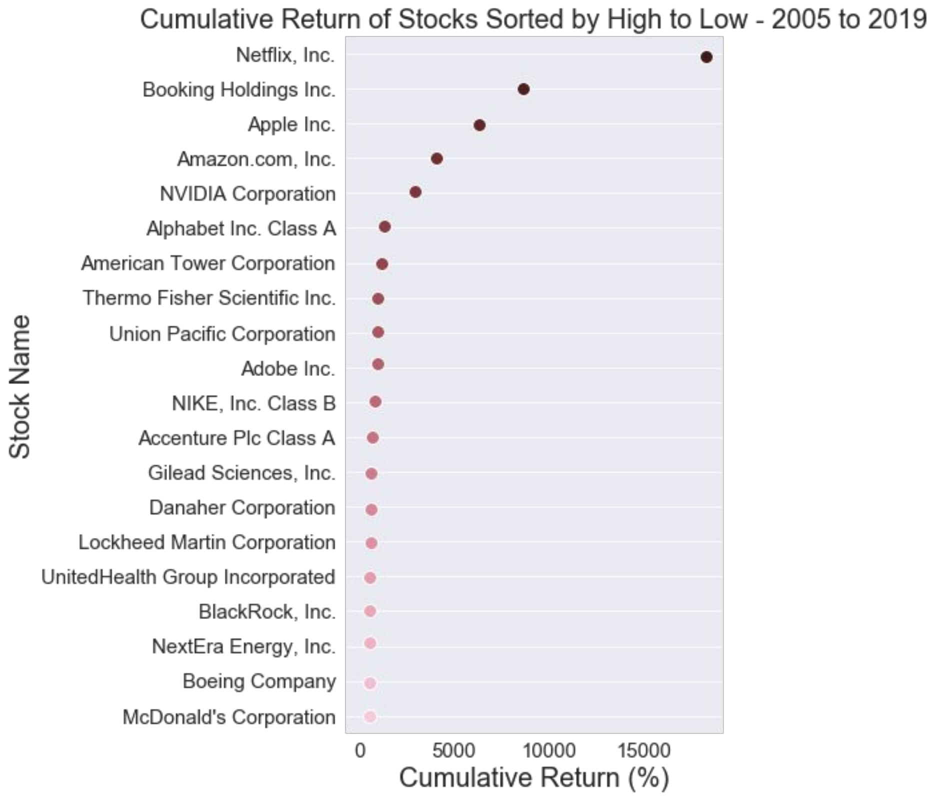


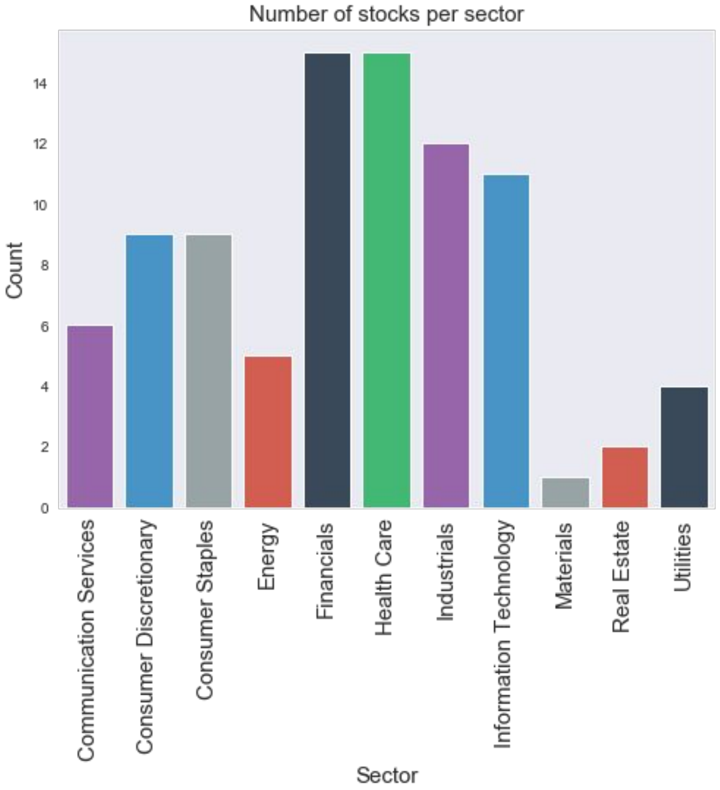
Exploratory Data Analysis - factors (part 2)





Exploratory Data Analysis - stocks





Results - ols

- Use Statsmodels package to run 89 regressions using OLS and append into an dictionary

```
ols_dict = {}
```

```
ols_dict
```

```
{}
```

```
for iden in np.unique(final_df.identifier):
    model = smf.ols(formula = 'excess_rets ~ mkt+ hml + smb + liq', data = final_df[final_df.identifier == iden]).fit()
    ols_dict.update([(iden, {'name':final_df[final_df.identifier == iden]['name'].max(),
                            'rsquared':model.rsquared,
                            'rsquared_adjusted':model.rsquared_adj,
                            'market':model.params['mkt'],
                            'liquidity':model.params['liq'],
                            'size':model.params['smb'],
                            'value':model.params['hml'],
                            'liquidity_pvalues':model.pvalues['liq'],
                            'sector':final_df[final_df.identifier == iden]['sector'].max(),
                            'cyc_or_def':final_df[final_df.identifier == iden]['cyclical_defensive'].max()})])
```

Results - pvalues

- Convert to dataframe and analyze how many of the regressions had a p-value of less than 0.05 for the liquidity factor

```
ols_df = pd.DataFrame.from_dict(ols_dict, orient = 'index')
```

```
ols_df.head()
```

	name	rsquared	rsquared_adjusted	market	liquidity	size	value	liquidity_pvalues	sector	cyc_or_def
AAPL-US	Apple Inc.	0.368554	0.354121	1.327989	0.526554	-0.555241	-0.626235	0.002884	Information Technology	Cyclical
ABT-US	Abbott Laboratories	0.243374	0.226079	0.654364	-0.189335	-0.305934	-0.402946	0.070267	Health Care	Defensive
ACN-US	Accenture Plc Class A	0.464867	0.452636	0.960305	-0.010841	0.069490	-0.414001	0.913175	Information Technology	Cyclical
ADBE-US	Adobe Inc.	0.488339	0.476644	1.457942	0.031238	-0.015479	-0.210736	0.828055	Information Technology	Cyclical
AGN-US	Allergan plc	0.267941	0.251208	0.904011	-0.529420	0.224308	-0.856511	0.000771	Health Care	Defensive

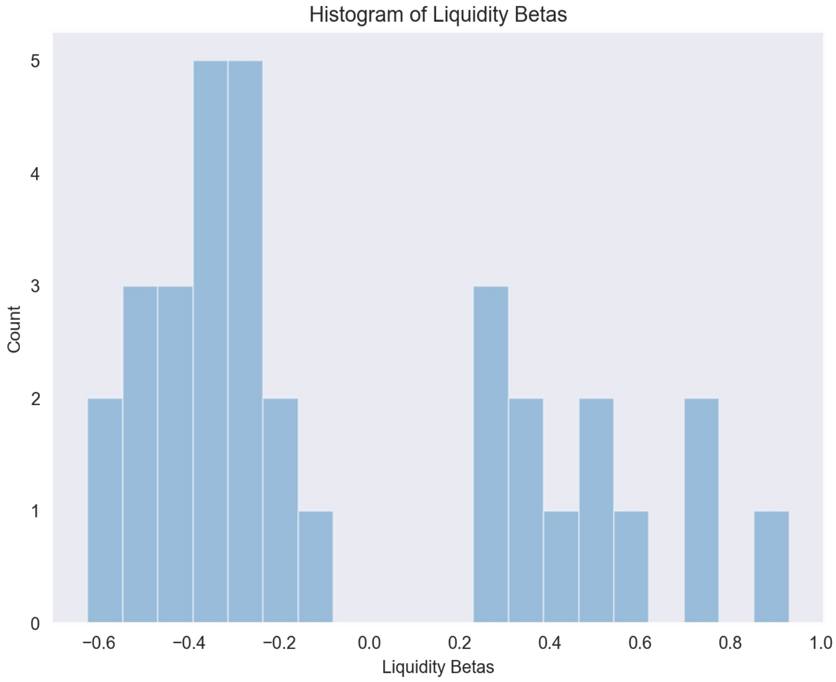
```
liquidity_df = ols_df[ols_df['liquidity_pvalues'] < 0.05]
```

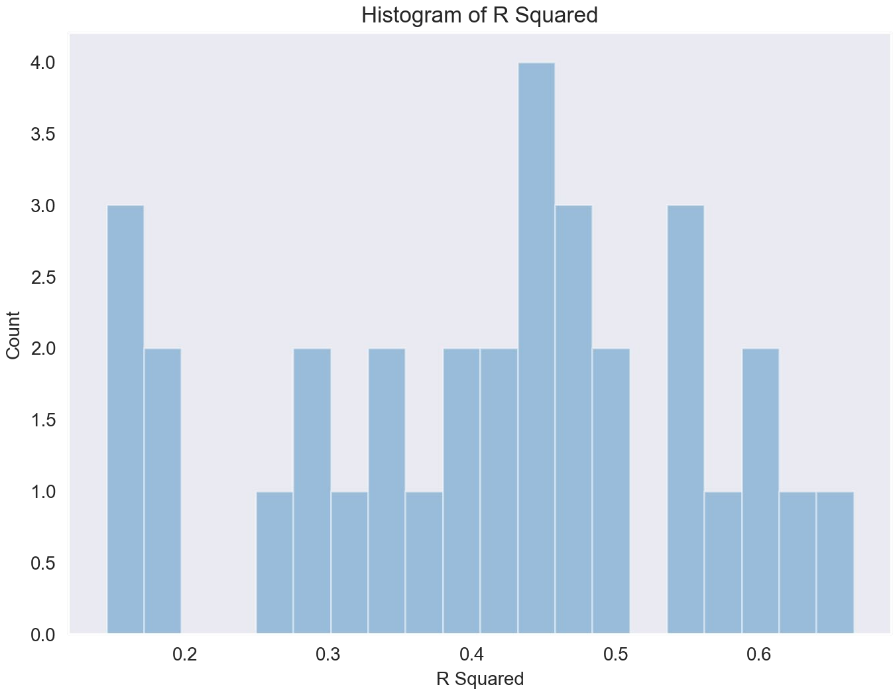
```
liquidity_df.shape
```

```
(33, 10)
```

Results - distribution

- Digging deeper we can see the distribution of the liquidity betas and R Squared

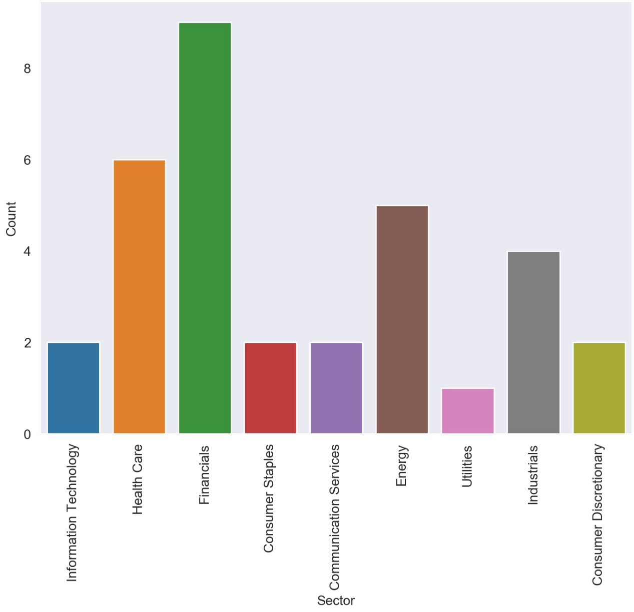




Results - inference

- Which sectors are influenced by liquidity the most

Barplot to show sectors that could be more influenced by liquidity

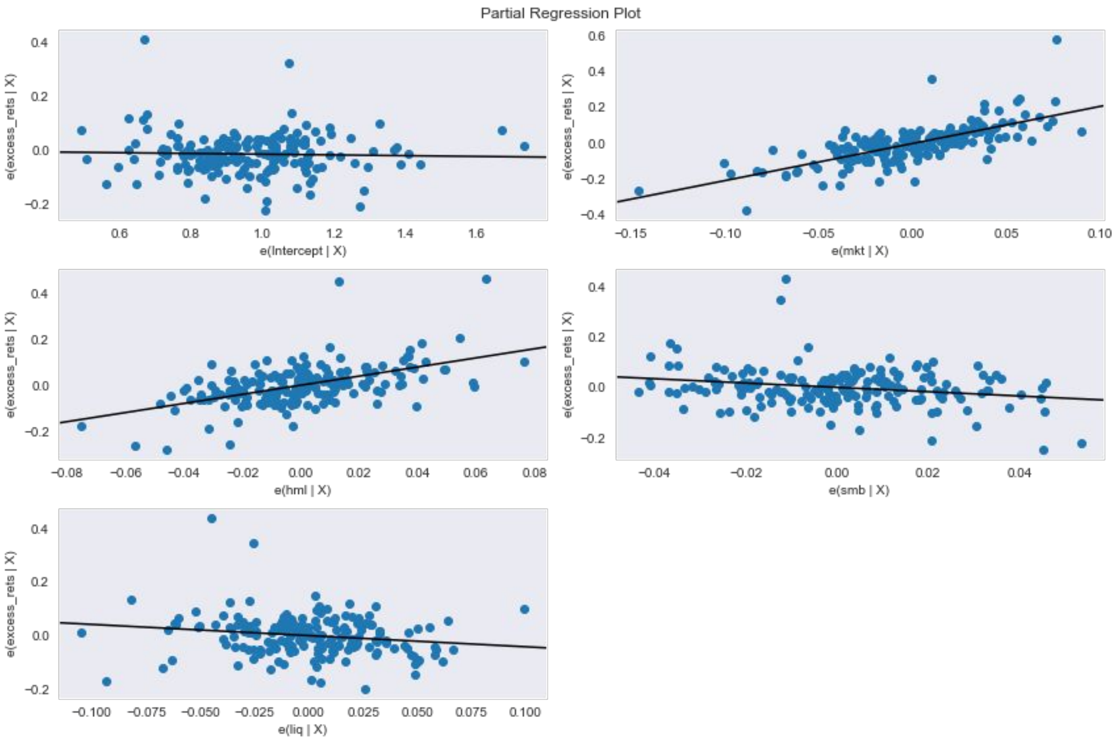


- The stock with the highest R Squared is Citigroup which is a US bank

```
liquidity_df.loc[liquidity_df['rsquared'].idxmax()]
```

```
name                     Citigroup Inc.
rsquared                       0.665793
rsquared_adjusted              0.658154
market                          2.06724
liquidity                     -0.420138
size                          -0.859033
value                           1.97508
liquidity_pvalues             0.0197262
sector                       Financials
cyc_or_def                     Cyclical
Name: C-US, dtype: object
```

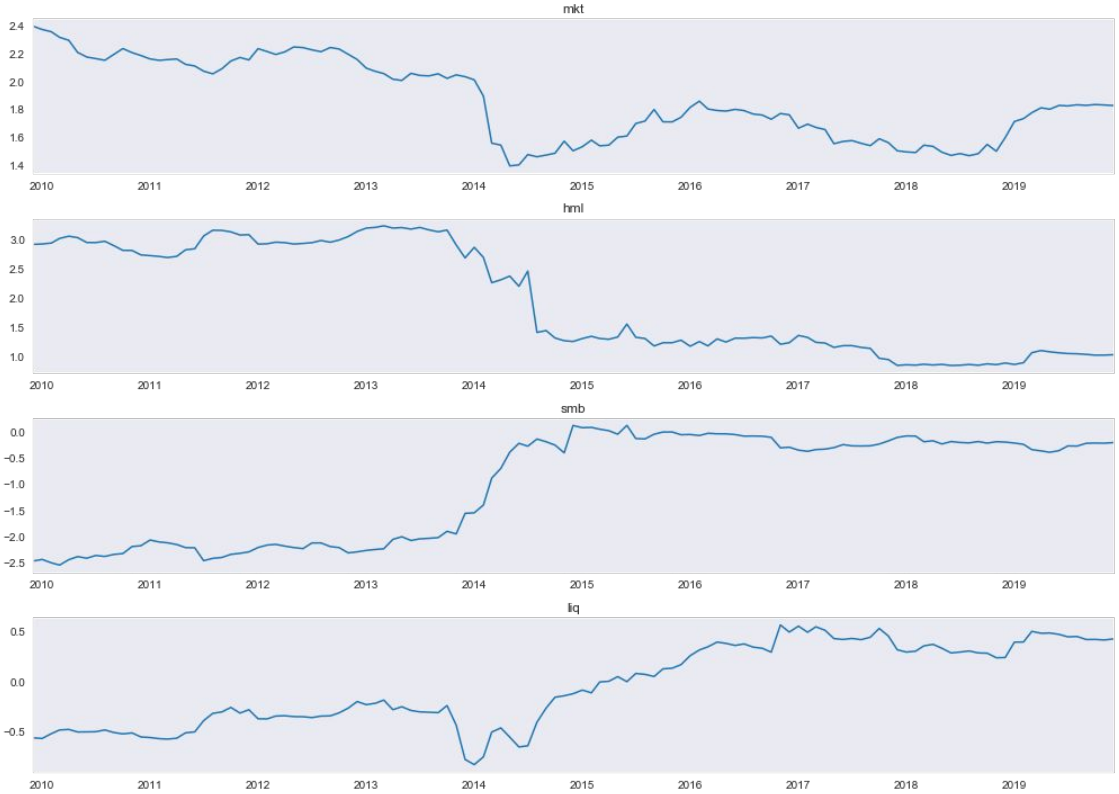
Results - partial regressions



- For Citigroup, partial regression plots show us that liquidity is an important factor for stock return but not as important as the market

- This can be seen by the strength of the slope

Results - rolling regressions



- For Citigroup, a rolling regression shows the liquidity beta to vary over time
- For time series data this is important as financial markets are very dynamic!
- Stocks change over time and a factor which is important in one time period may not be in another

Conclusion

- 33 stocks had liquidity as a significant factor with a p-value less than 0.05 out of 89 stocks
- A high proportion of these stocks were financials
- Only 8 stocks had a R Squared over 50% for the OLS model
- Things to improve
  - Using log returns
  - Use portfolio returns (less noisy than stock returns)
  - Perhaps narrow the number of stocks but use a longer time horizon
  - Testing the assumptions for regression (e.g. normality, homoscedasticity, autocorrelation)
  - Fine tuning the number of independent variables in the model
  - More detailed literature review
  - Using machine learning techniques such as time series split

Thanks for listening

Any questions?