A Comparison Between Different Factor Models*

Xuntian Lin, Sue Lu, Aman Halwasia
AFM 423 Machine Learning Approach to Quantitative Investing
School of Account and Finance
University of Waterloo

April 24, 2023

*We thank to Professor Wirjanto and Kelly Qin.

A Comparison Between Different Factor Models

Abstract

The paper's primary goal is to assess the predictive precision of two liquidity-based three-factor models, SiLiq and DiLiq, which were created as possible enhancements to the Fama-French model. Using a portfolio of stocks as the test data and Fama French three factors to construct a new factor called LMH to assess the two liquidity based modes.

1 Research Questions

Liquidity is a complex and multifaceted concept that encompasses the ability to trade quickly, at low cost, and with minimal price impact. According to Liu (2004), there are four dimensions of liquidity: trading quantity, trading speed, trading costs, and price impact. However, none of the proposed proxies have been able to fully capture all of these dimensions. While the bid-ask spread of Amihud and Mendelson (1986) is a direct measure of liquidity, it only focuses on trading costs. Compounding the issue is the challenge of obtaining sufficient data on bid-ask and other direct liquidity measures, which has hindered the integration of liquidity in asset pricing studies. To address this challenge, researchers such as Datar, Y. Naik, and Radcliffe (1998), and Amihud (2002) have resorted to alternative liquidity measures based on trading-volume variables. Hence we have the following research question waiting to be experimented and answered

1. How does the introduction of different factors will affect the Fama MacBeth regression results in comparison to the original Fama French 3 factor model in predicting portfolio returns?

2 Variables and Measures

2.1 Dependent and Explainatory Variables

This study considers the dependent variables as the monthly value-weighted average rate of returns on the test portfolio.

This study also includes explanatory factors from two sources, namely the Fama-French 3 factor model factors and the illiquidity risk premium proposed in this research. They include the following

- Market Risk-free rate (`MKT_RF`)
- Size (`SMB`)
- Value (`HML`)
- Liquidity risk premium (`LMH`)

The variables MKR_RF, SMB, and HML can be retrieved from the Fama French 3-factor model while we will create a new liquidity based premium called LMH. This can be done by sorting the test portfolio into three ranks based on the Share Turnovers from low to high for each month. The monthly value-weighted average rate of returns will be calculated for the three portfolios and LMH is obtained by subtracting the simple average of the returns on Low TURN portfolios from the simple average of the returns on High TURN portfolios.

The Share Turnovers (TURN), which is defined

$$\mathrm{TURN}_{j,t} = \frac{\mathrm{VOL}_{j,t}}{\mathrm{SHROUT}_{j,t}} \tag{1}$$

This is provided in the original database, where VOL is the trading volume and SHROUT is the number of shares outstanding.

3 The Application of the Machine Learning Approach to Factor Investing

3.1 CAPM

The Capital Asset Pricing Model (CAPM) is a broadly-used financial model that explains the relationship between the risk and return of an investment. It was developed by William Sharpe, John Lintner, and Jack Treynor in the 1960s.

Fundamentally, the CAPM suggests that the expected return of an investment is equal to the risk-free rate plus a risk premium, where the risk premium is proportional to the level of risk inherent in the investment. This risk premium is calculated by multiplying the investment's beta (a measure of its volatility relative to the market as a whole) by the market risk premium (the difference between the expected return of the market as a whole and the risk-free rate).

$$E_{R_i} = R_f + \beta_i(R_m - R_f) \tag{2}$$

The CAPM is often used as a tool for evaluating the performance of investment portfolios, as well as for determining the cost of capital for a company. While it has been criticized for its assumptions and limitations, it remains a useful and widely-used model in finance.

3.2 Fama French Three Factor Model

The Fama-French Three Factor Model is a financial model developed by Eugene Fama and Kenneth French in the 1990s. The model tries to explain the returns of a portfolio or an individual stock by incorporating three factors: market risk, size, and value. As of today, it has become a widely-used tool in the investment industry and has been used to develop investment strategies that attempt to outperform the market FAMA and FRENCH (1992).

$$E_{R_i} = R_f + \beta_1(R_m - R_f) + \beta_2(SMB) + \beta_3(HML) + \epsilon_{it} \tag{3}$$

The market risk factor, which is represented by the excess return of the market, reflects the overall risk of the stock market. The size factor, which measures the difference in returns between small and large companies, captures the fact that small companies tend to outperform large companies over time. The value factor, which measures the difference in returns between high and low book-to-market companies, captures the tendency for value stocks to outperform growth stocks over time.

3.3 Fama MacBeth Regression

This application is inspired by Fama MacBeth regression to assess the risk exposures. Asset pricing theories commonly use "risk factors" to explain asset returns, which can be macroeconomic (such as consumer inflation or unemployment rate) or financial (such as firm size, etc.). The Fama-MacBeth two-step regression is a practical method to assess how well these factors describe portfolio or asset returns and to determine the premium from exposure to these factors. In the first step, each portfolio's return is regressed against one or more factor time series to determine its exposure to each factor (the "factor exposures"). In the second step, the cross-section of portfolio returns is regressed against the factor exposures at each time step, producing a time series of risk premia coefficients for each factor. The key insight of Fama-MacBeth is to average these coefficients, once for each factor, to derive the premium expected for a unit exposure to each risk factor over time Scheuch, Voigt, and Weiss (2023).

In equation form, we have the following Fama MacBeth regression

$$
\begin{aligned}
R_{1,t} =& \alpha_1 + \beta_{1,F_1} F_{1,t} + \beta_{1,F_2} F_{2,t} + \ldots + \beta_{1,F_m} F_{m,t} + \epsilon_{1,t} \\
R_{2,t} =& \alpha_2 + \beta_{2,F_1} F_{1,t} + \beta_{2,F_2} F_{2,t} + \ldots + \beta_{2,F_m} F_{m,t} + \epsilon_{2,t} \\
&\vdots \\
R_{n,t} =& \alpha_n + \beta_{n,F_1} F_{1,t} + \beta_{n,F_2} F_{2,t} + \ldots + \beta_{n,F_m} F_{m,t} + \epsilon_{n,t}
\end{aligned} \tag{4}
$$

At each time period t, $R_{i,t}$ represents the return of portfolio or asset i (n total), while $F_{j,t}$ represents factor j (m total). The factor exposures, or loadings, denoted by β_{i,F_m}, describe how the returns are influenced by the factors. The value of t ranges from 1 through T. It is important to note that each regression employs the same set of factors F, as the aim is to ascertain the exposure of each portfolio's return to the given set of factors.

In the second step, the aim is to determine the exposure of the n returns to the m factor loadings over time, which is achieved by computing T cross-sectional regressions of the returns on the m estimates of the βs, denoted as $\hat{\beta}$, that were calculated in the first step. It is important to note that each regression employs the same βs from the first step because the objective now is to investigate whether a higher factor exposure corresponds to a higher return.

$$
\begin{aligned}
R_{i,1} =& \gamma_{1,0} + \gamma_{1,1}\hat{\beta}_{i,F_1} + \gamma_{1,2}\hat{\beta}_{i,F_2} + \ldots + \gamma_{1,m}\hat{\beta}_{i,F_m} + \varepsilon_{i,1} \\
R_{i,2} =& \gamma_{2,0} + \gamma_{2,1}\hat{\beta}_{i,F_1} + \gamma_{2,2}\hat{\beta}_{i,F_2} + \ldots + \gamma_{2,m}\hat{\beta}_{i,F_m} + \varepsilon_{i,2} \\
&\vdots \\
R_{i,T} =& \gamma_{T,0} + \gamma_{n,1}\hat{\beta}_{i,F_1} + \gamma_{n,2}\hat{\beta}_{i,F_2} + \ldots + \gamma_{n,m}\hat{\beta}_{i,F_m} + \varepsilon_{i,T},
\end{aligned} \tag{5}
$$

In each regression, the coefficients γ are calculated using the returns R, which are the same as those in (4). These coefficients are then used to determine the risk premium for each factor, and for each regression i varies from 1 through n.

The final output of the Fama-MacBeth regression is a set of $m+1$ series γ, which includes the constant term from the second step, for each factor. These series are of length T. Assuming the error terms ε are independently and identically distributed, the risk premium γ_m for factor Fm can be calculated by averaging the mth γ series over T. Additionally, standard deviations and t-stats can be obtained, including the t-stats for the mth risk premium:

$$\frac{\gamma_m}{\sigma_{\gamma_m}/\sqrt{T}} \tag{6}$$

Essentially, we will fun Fama MacBeth regression on the Fama French three factors, in additional to another liquidity based factor called LMH, which was referenced earlier in section 2.

4 Experimental Methodology

4.1 Dataset Preparation

This experiment dataset comprises information on 1,207 stocks listed in the US (N=1,207). The time period starts from November 11, 1998 to March 31, 2019 for a total of 244 monthly observations (T=244). For each point in time, there are 93 characteristics (K=93) of firms in the sample associated with each of the 1,207 stocks. Another database used to construct the Fama MacBeth regression is the monthly Fama French 3 factors From July 1926 to February 2023.

As software, we will use R code to import all the necessary datasets and essentialy libraries including `glmnet, dplyr, tidyverse, lubridate, quantmod, xtable` as the requirements.

4.2 Fama French Three Factor

When the Fama French three factor data is imported, we will merge the original stock dataset with the market risk-free, Size and value factors. Built-in ggplot functions was used to visualize the average monthly return aggregated over each calendar year for the three factors.

4.3 Creation of a New Factor LMH

To create the new factor LMH, avergae turn over is used to calculate the LMH (1). However, three options are available and they are

1. Average share turnover 3 months (`Share_Turn_3M`)
2. Average share turnover 6 months (`Share_Turn_6M`)
3. Avergae share turnover 12 months (`Share_Turn_12M`)

We will run the regression in three trials with the three options available above and comapre their factor loadings after Fama MacBeth regression

4.4 Fitting into Regressions

The first model we are going to fit is the original Fama French three factor model and run the Fama MacBeth Regression.

$$R_{i,t} = \alpha_i + \beta_i \left(R_{F,t}\right) + s_i \left(SMB_t\right) + d_i \left(HML_t\right) + \varepsilon_{i,t} \tag{7}$$

The variables in this model include realized returns (R_i) on the test portfolios, with α_i as the intercept, b_i, s_i, and d_i representing the estimated factor loadings. The market portfolio's realized returns (RM) and the market returns (RF) are also included. Finally, $\epsilon_{i,t}$ represents the error term at the end of month t.

To introduce the LMH factor, we use the strategy in paper Nor (1970) and introduce two new regression models

$$R_{i,t} = \alpha_i + \beta_i \left(R_{F,t}\right) + s_i \left(SMB_t\right) + l_i \left(LMH_t\right) + \varepsilon_{i,t} \tag{8}$$

The liquidity-based models proposed retain market risk as the primary risk factor but remove HML. The first model variant, called "SiLiq," combines the market return (RF) with premiums for "Size" (SMB) and "Liquidity" (H).

The second variant, known as "DiLiq," eliminates the size premium (SMB) and instead combines the market returns (MKT_RF), "DIstress" (HML), and LIQuidity (H) premiums. The regrssion is defined as follows

$$R_{i,t} = \alpha_i + \beta_i \left(R_{F,t}\right) + h_i \left(HML_t\right) + l_i \left(LMH_t\right) + \varepsilon_{i,t} \tag{9}$$

where α_i is the intercept, β_i, h_i, l_i and representing the estimated factor loadings.

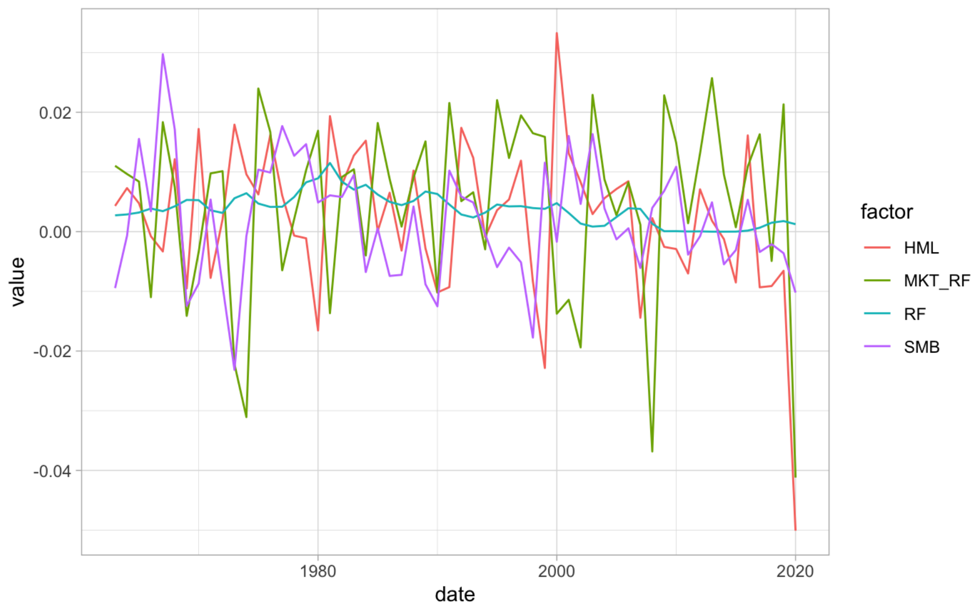
4.5 Calculation and Plotting

After the second stage of the Fama MacBeth Regression, we will plot the estimated factor premiums for each of the factor over the date. Mean value of the premiums will be calculated in addition to their t-statistics.

5 Results and Discussions

5.1 Fama French Three Factor

Figure 1. Comparison of the four factors from Fama French three factors



From Figure 1, comparing the different factors and their returns it is apparent that the market is the riskiest investment showing the highest volatility. This aligns with the high returns made in short periods of time. It should be noted that over the years the size factor SMB has become a lot less responsive to changes in the market rate. This could indicate a trend towards size no longer being a factor that corresponds well towards returns in a market. This makes sense as in the past a company simply being mature meant it had good growth and management however with an ever increasing number of large cap firms size is no longer a good representation of market returns.

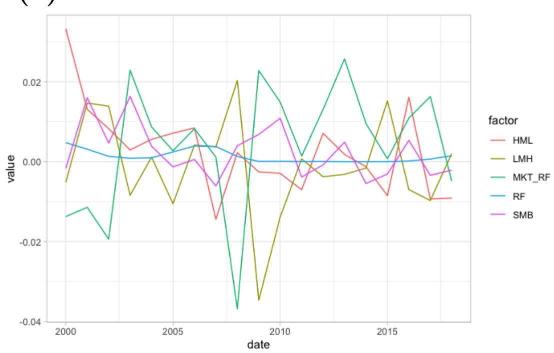
5.2 Fama French Three Factor Plus LMH

In Figure 2, we visualize the aggregated monthly return for the four factors including the LMH factor newly created for the purpose of this study. Comparing the 3 different options to create the LMH factor for the model. It is apparent that there is no major difference between the 3, 6 and 12 month share turnover. However it is important to note that between the 3 datasets the one closest to mirroring the data of the market risk free rate would be that of 12 month share turnover. This corresponds with a qualitative approach of a larger time frame providing enough time to smooth out any shocks to the market and is more accurate in representing the larger share of the market. Hence based on the graphs the 12 month share turnover is the most accurate.

Figure 2. Factor Trends

(a) LMH factor based on Share_Turn_3M

(b) LMH factor based on Share_Turn_6M



(c) LMH factor based on Share_Turn_12M

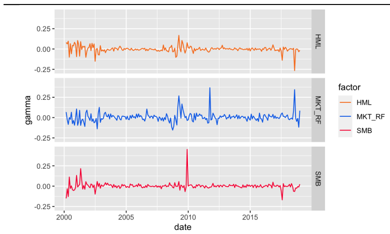
5.3 Descriptive statistics and Correlation

Begin with Figure 3, we have a table of different factors premiums based on different regression models. For Panel A, as the LMH factor is calculated on the top of Share_Turn_3M. Among the three different models, all the factors present volatile charateristics from 1998 to 2019. Panel B and Panel C show similar trends as the Panel A that all the factor premiums are volatile with major spikes observed during some specific times (e.g. 2008).

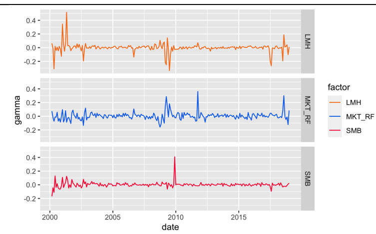
Next we computed the descriptive statistics and the correlation coefficients of the factors with relation to the share turnover, liquidity. Using this we were able to understand the relationship

Figure 3. Factor Risk Premiums Time Series

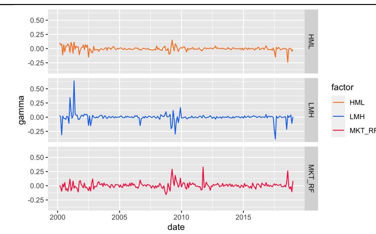
Panel A. Regression model based on variable Share Turn 3M



(a) Normal Model

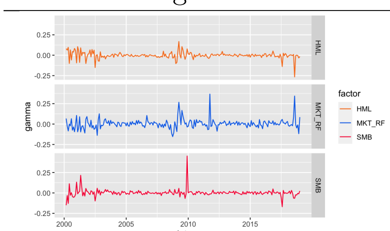


(b) SiLiq Model

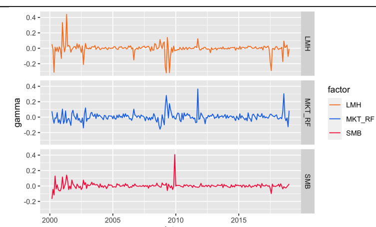


(c) DiLiq Model

Panel B. Regression model based on variable Share Turn 6M



(a) Normal Model

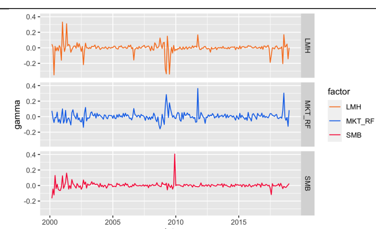


(b) SiLiq Model

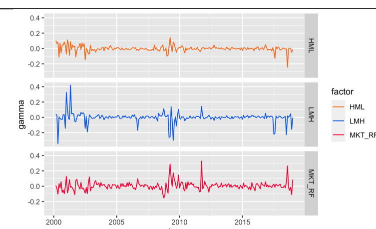
(c) DiLiq Model

Panel C. Regression model based on variable Share Turn 12M

(a) Normal Model



(b) SiLiq Model



(c) DiLiq Model

between the factors. After assessing the 3 tables it is clear that there are no significant changes across the share turnover durations as the t-stats stay consistent across them all. One change to be noted that follows along our previous explanation is the reducing correlation between the variables as the share turnover duration increased. This makes sense as over a larger period of time there is a greater possibility for other factors to influence the market price reducing the impact of liquidity as a factor.

Table 1. Descriptive statistics and correlation coefficients based on the variable Share_Turn_3M

Panel A. Explanatory factors in Fama-French Model						
	Mean	T-statistics	Std. Dev.	MKT_RF	SMB	HML
MKT_RF	0.005	1.204	0.059	1.00		
SMB	0.003	0.940	0.048	0.002	1.00	
HML	−0.002	−0.889	0.042	−0.199	−0.003	1.00
Panel B. Explanatory factors in SiLiq						
	Mean	T-statistics	Std. Dev.	MKT_RF	SMB	HML
MKT_RF	0.004	1.155	0.058	1.00		
SMB	0.004	1.205	0.045	0.043	1.00	
LMH	−0.002	0.400	0.070	−0.084	0.111	1.00
Panel C. Explanatory factors in DiLiq						
	Mean	T-statistics	Std. Dev.	MKT_RF	HML	LMH
MKT_RF	0.005	1.459	0.057	1.00		
HML	−0.003	−1.051	0.042	−0.149	1.00	
LMH	0.002	0.402	0.075	−0.084	0.144	1.00

Table 2. Descriptive statistics and correlation coefficients based on the variable Share_Turn_6M

Panel A. Explanatory factors in Fama-French Model						
	Mean	T-statistics	Std. Dev.	MKT_RF	SMB	HML
MKT_RF	0.005	1.204	0.059	1.00		
SMB	0.003	0.940	0.048	0.002	1.00	
HML	−0.002	−0.889	0.042	−0.199	−0.003	1.00
Panel B. Explanatory factors in SiLiq						
	Mean	T-statistics	Std. Dev.	MKT_RF	SMB	HML
MKT_RF	0.004	1.141	0.058	1.00		
SMB	0.004	1.216	0.045	0.048	1.00	
LMH	−0.003	−0.631	0.069	−0.088	0.130	1.00
Panel C. Explanatory factors in DiLiq						
	Mean	T-statistics	Std. Dev.	MKT_RF	HML	LMH
MKT_RF	0.005	1.455	0.057	1.00		
HML	−0.003	−1.083	0.042	−0.147	1.00	
LMH	−0.003	−0.646	0.073	−0.079	0.151	1.00

Table 3. Descriptive statistics and correlation coefficients based on the variable Share_Turn_12M

Panel A. Explanatory factors in Fama-French Model						
	Mean	T-statistics	Std. Dev.	MKT_RF	SMB	HML
MKT_RF	0.005	1.204	0.059	1.00		
SMB	0.003	0.940	0.048	0.002	1.00	
HML	−0.002	−0.889	0.042	−0.199	−0.003	1.00
Panel B. Explanatory factors in SiLiq						
	Mean	T-statistics	Std. Dev.	MKT_RF	SMB	HML
MKT_RF	0.004	1.147	0.058	1.00		
SMB	0.004	1.179	0.046	0.041	1.00	
LMH	−0.003	−0.622	0.068	−0.053	0.101	1.00
Panel C. Explanatory factors in DiLiq						
	Mean	T-statistics	Std. Dev.	MKT_RF	HML	LMH
MKT_RF	0.005	1.453	0.057	1.00		
HML	−0.003	−1.091	0.042	−0.142	1.00	
LMH	−0.003	−0.720	0.070	−0.059	0.094	1.00

Comparing our results to those of the previous study our results contradict the previous study to suggest that no such relationship that is large and positive exists for any of the factors in the competing models, including SMB. However the table makes it clear that CAPM is not sufficient in predicting market returns as there are several factors that are being missed outside of market risk. Although liquidity has not been statistically significant in outperforming CAPM, our findings echo previous iterations in suggesting that they still perform considerably better than other iterations of the 3 factor models.

In comparison to the original methodologies reported in Nor (1970), the numbers and statistics calculated from Table 1, 2, and 3 are aligned with the numbers obtained from the orginal paper.

6 Related Work

Projects 1 and 2 are highly related, therefore the reader is referred to our GPR #1 report for related work. In GPR #1, readers can refer to the paper for more information regarding the strengths of liquidity based models and possible improvements on those models.

The source code for this project is uploaded to the GitHub: `https://github.com/apokali/AFM-423`.

7 Conclusions

The initial findings of the study confirm the prevailing view that multifactor models outperform the CAPM in predicting stock returns. The study clearly demonstrates that the market factor (RM − RF) alone is insufficient to capture all the risks associated with stocks. Consequently, investors in this equity market must consider not only the market factor but also firm-specific factors, such as distress and liquidity levels. Although the forecasting accuracy of the competing three-factor models is consistently insignificant, DiLiq slightly outperforms the others, partly due to its inclusion of distress (as proxied by HML) and illiquidity (as proxied by H) as additional risk factors. This finding aligns with the concerns of investors in emerging equity markets regarding liquidity. It also strengthens the argument that investors require additional premiums to compensate for risks associated with distress and illiquidity, rather than solely risks associated with being small. In other words, being small does not necessarily make a company riskier; it is the company's potential risk of being in distress and losing liquidity that causes investors to demand higher-than-market risk premiums.

In summary, our empirical results strongly endorse the current perspective on the importance of liquidity in asset pricing models. Moreover, the insignificant variations in the forecasting accuracy suggest that the liquidity-based models are consistent with the Fama-French model. The three different vatiations used to create the LMH factor does not throw a significant impact to the results of the regression results.

References

Amihud, Y. 2002. Illiquidity and stock returns: Cross-section and time-series effects. *Journal of Financial Markets* 5:31–56.

Amihud, Y., and H. Mendelson. 1986. Asset pricing and the bid-ask spread. *Journal of Financial Economics* 17:223–249.

Datar, V. T., N. Y. Naik, and R. Radcliffe. 1998. Liquidity and stock returns: An alternative test. *Journal of Financial Markets* 1:203–219.

FAMA, E. F., and K. R. FRENCH. 1992. The cross-section of expected stock returns. *The Journal of Finance* 47:427–465.

Liu, W. 2004. Liquidity premium and a two-factor model. *SSRN Electronic Journal* .

Nor, R. A. R. . A. H. S. M. 1970. A COMPARISON BETWEEN FAMA AND FRENCH MODEL AND LIQUIDITY-BASED THREE-FACTOR MODELS IN PREDICTING THE PORTFOLIO RETURNS. *Asian Academy of Management Journal of Accounting and Finance (AAMJAF)* URL `https://ideas.repec.org/a/usm/journl/aamjaf00202_43-60.html`.

Scheuch, C., S. Voigt, and P. Weiss. 2023. Fama-Macbeth regressions. *Tidy Finance with R* p. 127–132.