

# STAT 222 - Feature Selection

AUTHOR  
Kenneth Chen

## Recovery Stage

### Stepwise Selection - Low Return

The table below contains the following regression information for the following stepwise selections:

- 1. Backward AIC
- 2. Backward BIC
- 3. Forward AIC
- 4. Forward BIC

[1] "R^2 for Full Model: 0.773" [1] "Adjusted R^2 for Full Model: 0.756"

	Dependent variable:			
	R1-RF			
	(1)	(2)	(3)	(4)
CMA	0.259*** (0.061)	0.238*** (0.054)	0.016 (0.181)	0.238*** (0.054)
CRD	-0.049* (0.025)			
LIQ	-0.152*** (0.027)	-0.156*** (0.023)	-0.153*** (0.023)	-0.156*** (0.023)
ME	0.118*** (0.041)	0.140*** (0.040)	0.074 (0.053)	0.140*** (0.040)
Mkt-RF	0.581*** (0.028)	0.571*** (0.027)	0.612*** (0.039)	0.571*** (0.027)
RMW	-0.134** (0.056)		-0.180** (0.089)	
ROE	0.096* (0.050)			
PC2			-0.234 (0.157)	
Constant	0.250*** (0.052)	0.222*** (0.050)	0.236*** (0.051)	0.222*** (0.050)
Observations	175	175	175	175

R <sup>2</sup>	0.772	0.762	0.768	0.762
Adjusted R <sup>2</sup>	0.762	0.757	0.760	0.757
Residual Std. Error	0.626 (df = 167)	0.634 (df = 170)	0.629 (df = 168)	0.634 (df = 170)
F Statistic	80.654*** (df = 7; 167)	136.193*** (df = 4; 170)	92.747*** (df = 6; 168)	136.193*** (df = 4; 170)

Note:  $p < 0.1$ ;  **$p < 0.05$** ;  $p < 0.01$

[1] "Selected Factors: CMA, LIQ, ME, **Mkt-RF** "

## Stepwise Selection - Medium Return

The table below contains the following regression information for the following stepwise selections:

1. Backward AIC
2. Backward BIC
3. Forward AIC
4. Forward BIC

[1] "R<sup>2</sup> for Full Model: 0.747" [1] "Adjusted R<sup>2</sup> for Full Model: 0.728"

	Dependent variable:			
	R <sup>2</sup> -RF			
	(1)	(2)	(3)	(4)
CMA	0.221*** (0.061)	0.221*** (0.061)	0.221*** (0.061)	0.221*** (0.061)
LIQ	-0.131*** (0.026)	-0.131*** (0.026)	-0.131*** (0.026)	-0.131*** (0.026)
ME	0.154*** (0.045)	0.154*** (0.045)	0.154*** (0.045)	0.154*** (0.045)
<b>Mkt-RF</b>	0.593*** (0.031)	0.593*** (0.031)	0.593*** (0.031)	0.593*** (0.031)
Constant	0.467*** (0.057)	0.467*** (0.057)	0.467*** (0.057)	0.467*** (0.057)
Observations	175	175	175	175
R <sup>2</sup>	0.742	0.742	0.742	0.742
Adjusted R <sup>2</sup>	0.736	0.736	0.736	0.736
Residual Std. Error (df = 170)	0.716	0.716	0.716	0.716
F Statistic (df = 4; 170)	122.546***	122.546***	122.546***	122.546***

Note:  $p < 0.1$ ;  **$p < 0.05$** ;  $p < 0.01$

[1] "Selected Factors: CMA, LIQ, ME, **Mkt-RF** "

## Stepwise Selection - High Return

The table below contains the following regression information for the following stepwise selections:

1. Backward AIC
2. Backward BIC
3. Forward AIC
4. Forward BIC

[1] "R<sup>2</sup> for Full Model: 0.704" [1] "Adjusted R<sup>2</sup> for Full Model: 0.682"

	<i>Dependent variable:</i>			
	<b>R3-RF</b>			
	(1)	(2)	(3)	(4)
CMA	0.204*** (0.074)	0.204*** (0.074)	0.234*** (0.072)	0.289*** (0.070)
PC3			-0.204*** (0.064)	-0.157*** (0.056)
LIQ	-0.103*** (0.032)	-0.103*** (0.032)	-0.116*** (0.039)	
ME	0.176*** (0.055)	0.176*** (0.055)		
MOM			-0.065** (0.032)	
<b>Mkt-RF</b>	0.624*** (0.038)	0.624*** (0.038)	0.626*** (0.038)	0.583*** (0.031)
Constant	0.687*** (0.069)	0.687*** (0.069)	0.734*** (0.069)	0.701*** (0.069)
Observations	175	175	175	175
R <sup>2</sup>	0.700	0.700	0.700	0.685
Adjusted R <sup>2</sup>	0.693	0.693	0.691	0.679
Residual Std. Error	0.872 (df = 170)	0.872 (df = 170)	0.874 (df = 169)	0.891 (df = 171)
F Statistic	99.101*** (df = 4; 170)	99.101*** (df = 4; 170)	78.932*** (df = 5; 169)	123.674*** (df = 3; 171)

Note:

$p < 0.1$ ;  **$p < 0.05$** ;  $p < 0.01$

[1] "Selected Factors: CMA, LIQ, ME, **Mkt-RF** "

# Expansion Stage

## Stepwise Selection - Low Return

The table below contains the following regression information for the following stepwise selections:

- 1. Backward AIC
- 2. Backward BIC
- 3. Forward AIC
- 4. Forward BIC

[1] "R^2 for Full Model: 0.784" [1] "Adjusted R^2 for Full Model: 0.78"

	Dependent variable:			
	R1-RF			
	(1)	(2)	(3)	(4)
CMA	0.174** (0.083)		0.204*** (0.057)	0.267*** (0.053)
CRD	-0.045** (0.019)			
EG	0.077* (0.046)			
LIQ			-0.006 (0.022)	
HML	0.276*** (0.035)	0.309*** (0.026)	0.294*** (0.054)	0.071* (0.041)
IA	-0.146* (0.078)			
MOM	-0.064*** (0.021)	-0.084*** (0.019)		
PC5			-0.032 (0.097)	0.493*** (0.047)
PC7			0.039 (0.069)	0.398*** (0.050)
Mkt-RF	0.812*** (0.018)	0.797*** (0.018)	0.857*** (0.042)	0.590*** (0.020)
RMW	0.361*** (0.035)	0.352*** (0.032)	0.404*** (0.069)	
PC4			-0.204*** (0.047)	

SMB	0.337*** (0.028)	0.331*** (0.027)	0.305*** (0.043)	
Constant	0.056* (0.032)	0.060** (0.030)	0.029 (0.036)	0.147*** (0.031)
Observations	598	598	598	598
R <sup>2</sup>	0.783	0.778	0.783	0.760
Adjusted R <sup>2</sup>	0.780	0.776	0.780	0.758
Residual Std. Error	0.715 (df = 588)	0.721 (df = 592)	0.715 (df = 588)	0.749 (df = 592)
F Statistic	235.712*** (df = 9; 588)	414.416*** (df = 5; 592)	235.986*** (df = 9; 588)	375.036*** (df = 5; 592)

Note:  $p < 0.1$ ;  **$p < 0.05$** ;  $p < 0.01$

[1] "Selected Factors: HML, MOM, **Mkt-RF**, RMW, SMB"

## Stepwise Selection - Medium Return

The table below contains the following regression information for the following stepwise selections:

1. Backward AIC
2. Backward BIC
3. Forward AIC
4. Forward BIC

[1] "R<sup>2</sup> for Full Model: 0.781" [1] "Adjusted R<sup>2</sup> for Full Model: 0.777"

	Dependent variable:			
	<b>R<sup>2</sup>-RF</b>			
	(1)	(2)	(3)	(4)
CMA	0.176* (0.092)		0.185** (0.073)	0.302*** (0.046)
EG	0.120** (0.051)			
HML	0.325*** (0.037)	0.306*** (0.028)	0.131** (0.057)	
IA	-0.216** (0.087)			
LIQ	0.047** (0.021)	0.057*** (0.021)		
MOM	-0.042* (0.022)			

ME			-0.205*	
			(0.119)	
Mkt-RF	0.856*** (0.022)	0.836*** (0.021)	0.671*** (0.048)	0.609*** (0.036)
PC5			0.436*** (0.113)	0.579*** (0.075)
PC7			0.478*** (0.104)	0.571*** (0.079)
RMW	0.358*** (0.040)	0.381*** (0.036)	0.191*** (0.073)	0.183*** (0.061)
SMB	0.376*** (0.033)	0.350*** (0.031)	0.361*** (0.122)	0.094** (0.045)
CRD			0.117*** (0.034)	0.134*** (0.031)
ROE			-0.165*** (0.060)	-0.248*** (0.053)
Constant	0.195*** (0.035)	0.210*** (0.034)	0.286*** (0.037)	0.310*** (0.035)
Observations	598	598	598	598
R <sup>2</sup>	0.781	0.775	0.781	0.778
Adjusted R <sup>2</sup>	0.777	0.773	0.777	0.775
Residual Std. Error	0.795 (df = 588)	0.804 (df = 592)	0.795 (df = 587)	0.799 (df = 589)
F Statistic	232.687*** (df = 9; 588)	406.890*** (df = 5; 592)	209.513*** (df = 10; 587)	258.286*** (df = 8; 589)

Note:

$p < 0.1$ ;  **$p < 0.05$** ;  $p < 0.01$

[1] "Selected Factors: HML, LIQ, Mkt-RF, RMW, SMB"

## Stepwise Selection - High Return

The table below contains the following regression information for the following stepwise selections:

1. Backward AIC
2. Backward BIC
3. Forward AIC
4. Forward BIC

[1] "R<sup>2</sup> for Full Model: 0.765" [1] "Adjusted R<sup>2</sup> for Full Model: 0.76"

Dependent variable:

	R3-RF			
	(1)	(2)	(3)	(4)
CMA	0.158 (0.108)			
EG	0.176*** (0.060)	0.187*** (0.060)		
HML	0.350*** (0.044)	0.322*** (0.035)	0.076 (0.055)	0.193*** (0.038)
IA	-0.245** (0.104)			
MOM			0.290*** (0.067)	0.114*** (0.032)
ME			-0.239* (0.131)	-0.328** (0.128)
ROE			-0.371*** (0.123)	
LIQ	0.153*** (0.025)	0.162*** (0.024)	-0.120** (0.054)	
Mkt-RF	0.903*** (0.025)	0.903*** (0.025)	0.716*** (0.040)	0.790*** (0.028)
RMW	0.367*** (0.047)	0.356*** (0.045)		
PC5			0.902*** (0.155)	0.475*** (0.060)
SMB	0.402*** (0.038)	0.409*** (0.038)	0.267* (0.151)	0.516*** (0.126)
PC7			0.357*** (0.068)	0.307*** (0.064)
PC6			0.633*** (0.125)	0.343*** (0.056)
Constant	0.319*** (0.042)	0.313*** (0.042)	0.540*** (0.050)	0.453*** (0.040)
Observations	598	598	598	598
R <sup>2</sup>	0.765	0.762	0.763	0.760
Adjusted R <sup>2</sup>	0.761	0.760	0.759	0.756
Residual Std. Error	0.952 (df = 589)	0.955 (df = 591)	0.956 (df = 587)	0.962 (df = 589)
F Statistic	239.157*** (df = 8; 589)	315.505*** (df = 6; 591)	189.293*** (df = 10; 587)	232.618*** (df = 8; 589)

Note:

$p < 0.1$ ;  $p < 0.05$ ;  $p < 0.01$

[1] "Selected Factors: EG, HML, LIQ,  $Mkt-RF$ , RMW, SMB"

# Downturn Stage

## Stepwise Selection - Low Return

The table below contains the following regression information for the following stepwise selections:

- 1. Backward AIC
- 2. Backward BIC
- 3. Forward AIC
- 4. Forward BIC

[1] "R<sup>2</sup> for Full Model: 0.723" [1] "Adjusted R<sup>2</sup> for Full Model: 0.718"

	Dependent variable:			
	$R1-RF$			
	(1)	(2)	(3)	(4)
CRD	-0.108*** (0.022)	-0.108*** (0.022)	-0.064** (0.028)	-0.065** (0.028)
HML	-0.108*** (0.036)	-0.108*** (0.036)		
IA	0.204*** (0.045)	0.204*** (0.045)		
LIQ	-0.064*** (0.021)	-0.064*** (0.021)		
$Mkt-RF$	0.660*** (0.018)	0.660*** (0.018)	0.559*** (0.020)	0.568*** (0.019)
RMW	0.220*** (0.038)	0.220*** (0.038)	0.321*** (0.052)	0.349*** (0.050)
PC8			0.115* (0.059)	0.132** (0.058)
SMB	0.122*** (0.029)	0.122*** (0.029)	0.057* (0.029)	0.057* (0.029)
PC7			-0.434*** (0.093)	-0.374*** (0.087)
EG			-0.310*** (0.086)	-0.252*** (0.079)



ROE			0.088*	
			(0.049)	
Constant	0.101*** (0.032)	0.101*** (0.032)	0.139*** (0.032)	0.140*** (0.032)
Observations	668	668	668	668
R <sup>2</sup>	0.722	0.722	0.723	0.721
Adjusted R <sup>2</sup>	0.719	0.719	0.719	0.718
Residual Std. Error	0.813 (df = 660)	0.813 (df = 660)	0.812 (df = 659)	0.813 (df = 660)
F Statistic	244.342*** (df = 7; 660)	244.342*** (df = 7; 660)	214.746*** (df = 8; 659)	244.125*** (df = 7; 660)

Note:  $p < 0.1$ ;  **$p < 0.05$** ;  $p < 0.01$

[1] "Selected Factors: CRD, HML, IA, LIQ, **Mkt-RF**, RMW, SMB"

## Stepwise Selection - Medium Return

The table below contains the following regression information for the following stepwise selections:

1. Backward AIC
2. Backward BIC
3. Forward AIC
4. Forward BIC

[1] "R<sup>2</sup> for Full Model: 0.721" [1] "Adjusted R<sup>2</sup> for Full Model: 0.716"

	Dependent variable:			
	R2-RF			
	(1)	(2)	(3)	(4)
CRD	-0.045* (0.025)		-0.037 (0.023)	
EG	-0.082 (0.052)			
HML	-0.140*** (0.040)	-0.093** (0.036)		
IA	0.234*** (0.049)	0.247*** (0.049)	0.168* (0.087)	
LIQ	-0.065*** (0.023)	-0.066*** (0.023)		
ME	0.122*** (0.031)	0.134*** (0.031)	0.078*** (0.030)	

Mkt-RF	0.699*** (0.020)	0.704*** (0.020)	0.795*** (0.032)	0.808*** (0.031)
RMW	0.263*** (0.047)	0.220*** (0.041)	0.399*** (0.053)	0.366*** (0.053)
CMA			0.034 (0.099)	0.249*** (0.041)
PC5			0.302*** (0.068)	0.307*** (0.067)
Constant	0.243*** (0.035)	0.222*** (0.035)	0.204*** (0.036)	0.188*** (0.035)
Observations	668	668	668	668
R <sup>2</sup>	0.720	0.716	0.719	0.714
Adjusted R <sup>2</sup>	0.716	0.714	0.716	0.712
Residual Std. Error	0.883 (df = 659)	0.887 (df = 661)	0.883 (df = 660)	0.889 (df = 663)
F Statistic	211.536*** (df = 8; 659)	278.060*** (df = 6; 661)	241.568*** (df = 7; 660)	413.866*** (df = 4; 663)
Note:				$p < 0.1$ ; $p < 0.05$ ; $p < 0.01$

[1] "Selected Factors: Mkt-RF , RMW, CMA, PC5"

## Stepwise Selection - High Return

The table below contains the following regression information for the following stepwise selections:

1. Backward AIC
2. Backward BIC
3. Forward AIC
4. Forward BIC

[1] "R<sup>2</sup> for Full Model: 0.739" [1] "Adjusted R<sup>2</sup> for Full Model: 0.735"

	Dependent variable:			
	R3-RF			
	(1)	(2)	(3)	(4)
EG	-0.134** (0.055)		-0.162*** (0.055)	
HML	-0.141*** (0.044)	-0.112*** (0.043)	-0.082* (0.049)	
IA	0.186*** (0.057)	0.192*** (0.057)	0.127** (0.057)	

ME	0.150*** (0.034)	0.167*** (0.033)		
MOM	0.089*** (0.025)	0.081*** (0.025)		
Mkt-RF	0.762*** (0.021)	0.769*** (0.021)	0.840*** (0.039)	0.744*** (0.018)
PC3			-0.169*** (0.035)	-0.209*** (0.033)
RMW	0.277*** (0.051)	0.213*** (0.044)	0.309*** (0.065)	
PC5			0.210** (0.087)	
Constant	0.375*** (0.039)	0.361*** (0.039)	0.369*** (0.041)	0.388*** (0.039)
Observations	668	668	668	668
R <sup>2</sup>	0.738	0.736	0.739	0.727
Adjusted R <sup>2</sup>	0.735	0.734	0.736	0.726
Residual Std. Error	0.987 (df = 660)	0.990 (df = 661)	0.986 (df = 660)	1.004 (df = 665)
F Statistic	265.956*** (df = 7; 660)	307.077*** (df = 6; 661)	266.588*** (df = 7; 660)	884.675*** (df = 2; 665)

Note:  $p < 0.1$ ;  **$p < 0.05$** ;  $p < 0.01$

[1] "Selected Factors: Mkt-RF , PC3"

## Depression Stage

### Stepwise Selection - Low Return

The table below contains the following regression information for the following stepwise selections:

1. Backward AIC
2. Backward BIC
3. Forward AIC
4. Forward BIC

[1] "R<sup>2</sup> for Full Model: 0.909" [1] "Adjusted R<sup>2</sup> for Full Model: 0.905"

Dependent variable:				
	R1-RF			
	(1)	(2)	(3)	(4)

CMA	0.243*** (0.073)	0.229*** (0.073)		
EG	-0.292*** (0.079)	-0.281*** (0.079)	-0.330*** (0.083)	-0.307*** (0.083)
MOM	-0.102*** (0.031)	-0.075*** (0.028)		
PC5			0.351*** (0.097)	0.314*** (0.095)
Mkt-RF	0.926*** (0.026)	0.914*** (0.025)	0.957*** (0.027)	0.930*** (0.024)
ME			-0.396* (0.209)	-0.215 (0.186)
RMW	0.266*** (0.086)	0.356*** (0.071)	0.221** (0.097)	0.338*** (0.074)
ROE	0.169* (0.090)		0.171* (0.092)	
SMB	0.468*** (0.050)	0.458*** (0.050)	0.748*** (0.187)	0.585*** (0.165)
PC2			0.227*** (0.075)	0.176** (0.071)
Constant	0.147** (0.071)	0.159** (0.071)	0.186*** (0.072)	0.183** (0.072)
Observations	287	287	287	287
R <sup>2</sup>	0.907	0.906	0.908	0.907
Adjusted R <sup>2</sup>	0.905	0.904	0.906	0.905
Residual Std. Error	1.153 (df = 279)	1.158 (df = 280)	1.149 (df = 278)	1.154 (df = 279)
F Statistic	389.798*** (df = 7; 279)	450.185*** (df = 6; 280)	343.824*** (df = 8; 278)	389.078*** (df = 7; 279)

Note:

$p < 0.1$ ;  **$p < 0.05$** ;  $p < 0.01$

[1] "Selected Factors: CMA, EG, MOM, Mkt-RF, RMW, SMB"

## Stepwise Selection - Medium Return

The table below contains the following regression information for the following stepwise selections:

1. Backward AIC
2. Backward BIC
3. Forward AIC

#### 4. Forward BIC

[1] "R^2 for Full Model: 0.892" [1] "Adjusted R^2 for Full Model: 0.887"

	Dependent variable:			
	R2-RF			
	(1)	(2)	(3)	(4)
CMA	0.119 (0.082)			
EG	-0.301*** (0.088)	-0.335*** (0.085)	-0.390*** (0.090)	-0.329*** (0.087)
LIQ	0.067* (0.037)			
Mkt-RF	0.900*** (0.030)	0.901*** (0.024)	0.932*** (0.029)	0.899*** (0.025)
ME			-0.103 (0.174)	0.051 (0.163)
RMW	0.310*** (0.092)	0.450*** (0.064)	0.369*** (0.083)	0.444*** (0.067)
ROE	0.147* (0.087)		0.181* (0.093)	
PC5			0.194* (0.104)	
SMB	0.476*** (0.057)	0.495*** (0.053)	0.606*** (0.176)	0.446*** (0.164)
Constant	0.299*** (0.077)	0.320*** (0.076)	0.318*** (0.076)	0.318*** (0.077)
Observations	287	287	287	287
R <sup>2</sup>	0.891	0.888	0.891	0.888
Adjusted R <sup>2</sup>	0.888	0.887	0.888	0.886
Residual Std. Error	1.238 (df = 279)	1.246 (df = 282)	1.239 (df = 279)	1.248 (df = 281)
F Statistic	326.060*** (df = 7; 279)	561.018*** (df = 4; 282)	325.089*** (df = 7; 279)	447.399*** (df = 5; 281)

Note:

$p < 0.1$ ;  $p < 0.05$ ;  $p < 0.01$

[1] "Selected Factors: EG, Mkt-RF, RMW, SMB"

## Stepwise Selection - High Return

The table below contains the following regression information for the following stepwise selections:

1. Backward AIC
2. Backward BIC
3. Forward AIC
4. Forward BIC

[1] "R<sup>2</sup> for Full Model: 0.88" [1] "Adjusted R<sup>2</sup> for Full Model: 0.875"

	<i>Dependent variable:</i>			
	R3-RF			
	(1)	(2)	(3)	(4)
EG	-0.349*** (0.096)	-0.349*** (0.096)	-0.515*** (0.101)	-0.515*** (0.101)
LIQ	0.099*** (0.038)	0.099*** (0.038)		
CMA			0.318*** (0.082)	0.318*** (0.082)
Mkt-RF	0.898*** (0.033)	0.898*** (0.033)	0.785*** (0.034)	0.785*** (0.034)
RMW	0.323*** (0.090)	0.323*** (0.090)		
ROE	0.236** (0.094)	0.236** (0.094)		
ME			-0.027 (0.188)	-0.027 (0.188)
PC4			-0.840*** (0.123)	-0.840*** (0.123)
SMB	0.504*** (0.061)	0.504*** (0.061)	0.585*** (0.188)	0.585*** (0.188)
Constant	0.430*** (0.083)	0.430*** (0.083)	0.594*** (0.084)	0.594*** (0.084)
Observations	287	287	287	287
R <sup>2</sup>	0.879	0.879	0.879	0.879
Adjusted R <sup>2</sup>	0.877	0.877	0.877	0.877
Residual Std. Error (df = 280)	1.347	1.347	1.347	1.347
F Statistic (df = 6; 280)	340.484***	340.484***	340.379***	340.379***

Note:  $p < 0.1$ ;  **$p < 0.05$** ;  $p < 0.01$

[1] "Selected Factors: EG, LIQ, Mkt-RF, RMW, ROE, SMB"

