TaylorFit

Not your usual fit



What does Taylorfit do?

- Regression program that makes fitting MPR models easy.
 - Multivariable polynomial regression (MPR)
- Produces a polynomial describing the relationship between data
- Helps to predict non-obvious relationships
- Math-magic

Where Can it be Used?

- Demand forecasting
- Business analytics
- Market research
- Data mining

So, what did we do?

- Started off as a DOS program.
 - Would not even compile in modern GCC.
- Modernized
- Extended -- Added functionality
 - Graphs
 - New math
- Ended as a modern QT5 GUI
- Documented EVERYTHING

Before AKA - Things that were still new in 1994

TaylorFit - 1994





After

TaylorFit 3.0 - 2015

				TaylorFit	x	×
Eile Edit Export View Help						-
Method Eit Parameters	Model					=
Exponents 1	ND	NP	NF	TSS	SSE MSE RSQ RSQ-adj FSTAT	
Multiplicands 1	Fit 393	4	389	23821.4	6990.2 17.9697 0.706558 0.704295 234.161	
Dep. Column						
Time Series Enable Disable						
Lags						
Use Lagged Dependent Variable						
Yes O No	e				QCustomPlot Interaction Example	×
Mode	Current Model - Mod	el Number 4	Drift)	col		
Manual	1 -0.0429376	0	0.01	4	Taylorfit	
Sort Candidate Terms By	2 -0.00524127	0	0.01	5		
 Best T-Statistic 	3 45.6707	0	1.11215	0	New graph 0	<u>'</u>
Maximize F-Statistic	4 -0.392171	0	0.01	2	180	
Minimize MSE					2 150	
Minimize AIC						
Minimize BIC					60	
Best T-Statistic &	Constants Trees				12 18 24 30 36 42	1
Maximize F-Statistic	Coeff	t-Stat	Pr(t)	col	x Axis	
Best T-Statistic & Minimize MSE	1 0.000140139	0	0.01	3		
	2 -0.0416605	0	0.01	6	Controls	
Clear Model	3 0.747782	0	0.01	7	4 v y-axis	
Reset Method Defaults	4 1.33896	0	0.01	8	1 v x-axis	
Update Fit Parameters					Line Markers Col vs col Time Series	
Graphs				50 + •	Select the areas to drag and soon them individually. Dealback fields tables or byped true to actual argumption Left click on graphs or legend to select graphs. Right click for a popup menu to addreemove graphs and move the legend	
				1		

Cross Platform Compatibility



Single Codebase



Demonstration

Fit Data: ens/Sen: Test Data: ens/Sen: Load Data CI GM 1 0.539	ior Design/1 ior Design/1 lear Data SE 34.69	taylorfit/sh taylorfit/sh Clear All M	are/RETAIL HE are/RETAIL HE Create Plots Y	ADERS.csv ADERS.csv	Data M Browse Browse	Row From: 2 Row From: 2	Row To: Row To:	55 55	 First line contains header First line contains header
Fit Data: ens/Sen: Test Data: ens/Sen: Load Data Cl GM 1 0.539	ior Design/1 ior Design/1 lear Data SE 34.69	taylorfit/sh taylorfit/sh Clear All M	are/RETAIL HE are/RETAIL HE Create Plots Y	ADERS.csv ADERS.csv	Browse Browse	Row From: 2 Row From: 2	Row To: Row To:	55 55	✓ First line contains header
Test Data: ens/Sen Load Data Cl GM 1 0.539	ior Design/1 lear Data SE 34.69	Clear All	are/RETAIL HE Create Plots Y	ADERS.csv	Browse	Row From: 2	Row To:	55	Eirst line contains header
Load Data Cl GM 1 0.539	SE 34.69	Clear All	Create Plots						rist file contains fleader
GM 1 0.539	SE 34.69	M	Y						
1 0.539 3	34.69	0 E		C					
1 0.555	54.05	N	10	3	3				
3 8 590	10 7	0.0	10	1	1				
2 0.589	10.7	0.95	15	1	1				
3 0.539	/9.12	0.3	30	3	3				
4 0.488	18.31	0.2	4	3	2				
5 0.391	34.72	0.4	5	3	1				
6 0.428	40.72	0.22	10	3	3				
7 0.382	13.82	0.25	10	3	2				
8 0.447	22.17	0.36	5	3	3				
9 0.43	28.14	0.45	7	3	2 .				
10 0.477	21.74	0.6	20	2	3				
11 0.395	10.48	0.1	10	1	3				
12 0.528	32.03	0.25	5	3	3				
13 0.565	16.7	0.6	7	3	2				
14 0.458	30.8	0.3	6	3	3				
15 0.524	19.2	0.6	15	1	2				
16 0.493	37.62	0.22	30	2	3				
17 0 522	18 98	0.5	4	-	2				
	10.90	0.5	*	2	2				

We select RETAIL HEADERS.csv as our fitting data set. The program populates the testing data with the same file if another file is not provided. The program also automatically detects if there is a string header in the file and loads it into the chart. The user can select more specific data ranges if they wish

•				TaylorFit				
lash a d		- 4-1		Dutu				
	M	odel						
Fit Parameters	G	lobal Model Sta	tistics					
xponents 1								
ultiplicands 1								
ep. Column 1								
ime Series								
Enable 💿 Disable								
ags								
se Lagged Dependent Variable								
Yes No	с	urrent Model			0			
lode		Coeff	t-Stat	Pr(t)	col	lag	exp	SSE-add
Manual ᅌ								
ort Candidate Terms By								
Best T-Statistic								
Minimize MSE								
Minimize AIC								
Minimize BIC	- 11				0			
Best T-Statistic &	С	andidate Terms						
Best T-Statistic &		Coeff	t-Stat	Pr(t)	col	lag	exp	SSE-add
Minimize MSE	1	0.813782	65.5947	0.397595	7	0	1	0.161446
Clear Model	2	0.492167	60.7776	0.389439	0	0	0	0.187675
Reset Method Defaults	3	0.177656	27.0543	0.251172	6	0	1	0.895874
Compute Fit		0 192626	22 0022	0 210550	-	0	1	
	- 4	0.103030	23.0033	0.210559	5	0	1	1.12799

When we switch to the model tab we see several options to set up the fit. When we compute the fit the table of candidate terms appears sorted by the specified criteria.

				Data	Model						
	Model										
rameters	Global Model Stati	stics									
nents 1	ND	NP	NF	TSS	SSE	MSE	RSQ	RSQ-adj	FSTAT	AIC	BIC
blicands 1	Fit 54	3	51	0.187676	0.0692744	0.00135832	0.630882	0.616407	29.0558	-2.75589	-2.77075
Column 1											
Contra 1											
ble Disable											
agged Dependent Variable											
○ No											
inual											
st T-Statistic	Current Model - M	odel Number 3				•					
aximize F-Statistic	Coeff	t-Stat	Pr(t)	col	lag	exp			SSE-ad	đ	
nimize MSE	1 0.458239	8.08118	0.0001	7	0	1	0.161446				
nimize AIC nimize BIC	2 0.181038	5.11551	0.0001	0	0	0	0.103211				
st T–Statistic &	3 8, 129886	4.9984	0.0001	3	8	1	0.0692744				
iximize F-Statistic				1	-	-					
nimize MSE											
Clear Model											
Reset Method Defaults											
Compute Fit											
Create Plots											
	Candidate Terms					•					
	Coeff	t-Stat	Pr(t)	col	lag	exp			SSE-ad	ł	
	1 0.00101302	1.60692	0.0001	4	9	1	0.0658725				
	2 0.0105838	1.50316	0.0001	5	0	1	0.0662792				
		0.069934	0.0001	6	0	1	0.0679978				
	3 -0.00946219	-0.2000.34		1	-	-					
	3 -0.00946219	-0.900034	0.0001	0	0	1	0 060667				
	3 -0.00946219 4 0.040655	0.665021	0.0001	8	0	1	0.068667				

Double clicking terms adds them to our model. Here when we add all terms with a T Statistic greater than 2. This is a linear model. Note that our R Squared is .63. Let's see if we can do better.

				Taylo Data	rFit Model								
thod	Model												
it Parameters	Global Model Statis	tics											
kponents 1 2 -1	ND	NP	NF	TSS	SSE	MSE	RSQ	RSQ-adj	FSTAT	AIC	BIC		
Itiplicands 1 2	Fit 54	3	51	0.187676	0.0692744	0.00135832	0.630882	0.616407	29.0558	-2.75589	-2.77075		
. Column 1													
e Series													
nable 💿 Disable													
Lagged Dependent Variable													
es 🔿 No													
le													
Innual													
anuai													
Candidate Terms By						0							
st T-Statistic	Current Model - Mo	del Number 3			1			1					
aximize F-Statistic	Coeff	t-Stat	Pr(t)	col	lag	exp	col	lag	exp		SSE-add		
finimize AIC	1 0.458239	8.08118	0.0001	7	9	1	0	0	0	0.15798			
finimize BIC	2 0.181038	5.11551	0.0001	0	0	0	0	0	0	0.104819			
est T-Statistic &	3 0.129086	4,9984	0.0001	3	0	1	0	0	0	0.103211			
Maximize F-Statistic													
Animize MSE													
Clear Model													
Reset Method Defaults													
Compute Fit													
Create Blots													
create nots													
						0							
	Candidate Terms												
	Coeff	t-Stat	Pr(t)	col	lag	exp	col	lag	exp		SSE-add		
	1 -0.0268353	-4.39202	0.0001	3	0	-1	4	0	-1	0.0499887			
	2 -0.0207442	-3.34551	0.0001	4	0	-1	6	0	2	0.0566037			
	3 -0.0002369	-3.01003	0.0001	3	0	-1	8	0	-1	0.0586471			
	4 -0.291368	-2.93729	0.0001	4	0	-1	5	0	-1	0.0590799			
	5 -0.00807407	-2.71082	0.0001	4	0	-1	8	0	-1	0.0603976			
	5 0.000245002	2,68982	0.0001	4	0	1	5	0	2	0.0605173			

7 -0.000595822 -2.6062

-2.58416

-2.57452

8 -0.00025704

9 -0.0411696

0.0001

0.0001

0.0001

3

6

4

0

0

0

-1

2

-1

6

8

6

0

0

0

2

-1

1

0.0609892

0.0611124

0.061166

We change our exponents to 1 2 and -1 and our multiplicands to 1 and 2. This results in a polynomial model with interactions including ratio terms.

•				Taylo	rFit						
				Data	Model						
thod	Model										
t Parameters	Global Model Stati	stics									
ponents 1 2 -1	ND	NP	NF	TSS	SSE	MSE	RSQ	RSQ-adj	FSTAT	AIC	BIC
ultiplicands 1 2	Fit 54	6	48	0.187676	0.0331236	0.000690074	0.823506	0.805121	37.3274	-2.93888	-2.96862
p. Column 1											
me Series											
Enable 💿 Disable											
32											
e Lagged Dependent Variable											
Yes 🔿 No											
ode											
Manual											
rt Candidate Terms By	Current Madel N	adal Norshan 1.4									
Best T-Statistic	Current Model - W	t-Stat	De(t)	col	120	ex2	col	120	exp		bbe-322
Minimize MSE		40.0004	0.000000	-			0		0	0.42642	336-800
Minimize AIC	1 0.522181	12.3324	0.0109099	/	0	1	0	0	0	0.13643	
Minimize BIC	2 0.148905	5.73744	0.0001	0	0	0	0	0	0	0.0500734	
Best T-Statistic & Maximize F-Statistic	3 0.00567077	2.90645	0.0001	5	0	2	7	0	1	0.0378119	
Best T-Statistic &	4 -0.0639988	-7.89084	0.0001	3	0	-1	4	0	-1	0.0547769	
Minimize MSE	5 1.42685	5.32289	0.0001	2	0	-1	6	0	-1	0.0442748	
Clear Model	6 0.0122863	3.30206	0.0001	4	0	-1	8	0	-1	0.0367418	
Reset Method Defaults											
Compute Fit											
Create Plots											
	Candidate Terms										
	Coeff	t-Stat	Pr(t)	col	lag	exp	col	lag	exp		SSE-add
	1 -0.0010694	-1.74005	0.0001	7	0	-1	8	0	-1	0.0311189	
	2 -4.46965e-05	-1.62039	0.0001	2	0	1	8	0	-1	0.031371	
	3 -0.00159375	-1.56845	0.0001	8	0	-1	0	0	0	0.0314761	
	4 0.145419	1.4688	0.0001	3	0	2	8	0	1	0.0316699	
	5 -0.00229787	-1.40935	0.0001	7	0	1	8	0	-1	0.0317805	
	6 0.00664776	1.40563	0.0001	3	0	2	6	0	2	0.0317873	
	7 0.0181224	1.39915	0.0001	3	0	2	6	0	1	0.0317991	
	8 0.385087	1.3891	0.0001	3	0	2	8	0	2	0.0318173	
	0 -0 000502214	-1 29501	0 0001	6	0	1	0	0	-1	0 0310070	

Next we add all terms with tstat > 2.0 AND remove terms with t-stat < 2.0. We do this iteratively until there are no terms t>2 in the Candidate Terms table and none with t<2 in the Current Model table.

•				Taylo	rFit						
				Data	Model						
Marked.											
Sit Deservation	Model										
Fit Parameters	Global Model Statis										
	ND	NP	NF	155	SSE	MSE	RSQ	RSQ-adj	FSTAT	AIC	BIC
Multiplicands 1 2	Fit 54	6	48	0.187676	0.0331236	0.000690074	0.823506	0.805121	37.3274	-2.93888	-2.96862
Dep. Column 1											
Time Series											
Enable 💽 Disable											
Lags											
Use Lagged Dependent Variable											
Yes No											
Mode											
Manual											
Manual											
Sort Candidate Terms By						0					
Best T-Statistic	Current Model - Mo	del Number 14		1					1		
Maximize F-Statistic	Coeff	t-Stat	Pr(t)	col	lag	exp	col	lag	exp		SSE-add
Minimize MSE	1 0.522181	12.3324	0.0109699	7	0	1	0	0	0	0.13643	
Minimize BIC	2 0.148905	5.73744	0.0001	0	0	0	0	0	0	0.0500734	
Best T-Statistic &	3 0.00567077	2,90645	0.0001	5	0	2	7	0	1	0.0378119	
Maximize F-Statistic	A -0.0620088	7 90094	0.0001	2	0	-1	4	0	-1	0.0547760	
Minimize MSE	4 -0.0035908	-7.09004	0.0001	3	0	-1	*	0	-1	0.0347709	
Clear Model	5 1.42685	5.32289	0.0001	2	0	-1	6	0	-1	0.0442748	
Reset Method Defaults	6 0.0122863	3.30206	0.0001	4	0	-1	8	0	-1	0.0367418	
Compute Fit											
Create Plots											
create riots											
						0					
	Candidate Terms										605 - 11
	Coerr	t-stat	Pr(t)	coi	lag	exp	coi	Tag	exp		55E-add
	1 -0.0010694	-1.74005	0.0001	7	0	-1	8	0	-1	0.0311189	
	2 -4.46965e-05	-1.62039	0.0001	2	0	1	8	0	-1	0.031371	
	3 -0.00159375	-1.56845	0.0001	8	0	-1	0	0	0	0.0314761	
	4 0.145419	1.4688	0.0001	3	0	2	8	0	1	0.0316699	
	5 -0.00229787	-1.48935	0.0001	7	0	1	8	0	-1	0.0317805	
		1.10553		,		-			-	0.0017000	
	0 0.00004770	1.40303	0.0001	3	U	2	0	U	4	0.031/8/3	
	7 0.0181224	1.39915	0.0001	3	0	2	6	0	1	0.0317991	
	8 0.385087	1.3891	0.0001	3	0	2	8	0	2	0.0318173	
	9 -0.000582214	-1.28591	0.0001	6	0	1	8	0	-1	0.0319978	

Note that R Squared is now .82. Since the amount of error is proportional to 1-RSQ, this means adding polynomial interaction terms decreased the relative error from 0.37 to 0.18, cutting the error by more than half.



TaylorFit can also be used to create graphs of models. The user may choose any of the columns to plot on the x or y axis and may rename and delete graphs. Graphs can be exported as images and pdfs.

Who are the competitors?

- R
- STATA
- SAS
- SPSS

- General purpose
 - Not just modeling
- Difficult to learn
- Limited to linear modeling and other constrained approaches

How does it compare to others?

- Others designed for Design of Experiment.
- TaylorFit allows fitting of Nth degree polynomials
- TaylorFit selection of terms using stepwise regression.
- Has a pleasant GUI.

Why No Artificial Neural Network?

- Good for accurate fits
 - Prone to overfit
- Can describe complex relationships
 - But the structure must be determined a priori

Where Did We Stumble?

- Engine is confusing and has lots of unneeded abstractions.
- Difficulty supporting all major OSs.
- Coordinating code was tough at first but git made that easy.

TaylorFit 4.0?

- Time Series
- Probability calculations
- History capability
- Sensitivity analysis capability
- Swap out core engine
 - o <u>http://www.astro.umd.edu/~bjw/software/boycottnr.html</u>
- Update save format
 - (it's really temperamental right now)

Acknowledgements (1/2) (in no order)

- David Adam Perez
- Brandon Schur
- Chris Barry
- Tim Barbra
- Professor Vaccari
- Professor Burlick
- StackOverflow
- Neal
- Mamouns
- That corner store next to Mamouns
- Qt Developers
- Nick Bevacqua
- Xebra Corporation
- Adam Cotenoff
- Mark Fleres
- Marc Kaminsky
- Victoria Albright
- Bier Garten

- Qdoba
- Perl (the human)
- John Coltrane
- Professor Gabarró
- Shamwow Guy
- Erich Polixa
- Radhika Kasabwala
- ???
- Randall Suliga
- Dan Coullias
- Jess Berg
- Christian Benitez
- The fairy wand
- Blase Feeney
- Matt Coccia
- Fred Florio
- GCC / Clang teams
- Kevin

Acknowledgements (2/2) (cont.)

- Bacon Pancakes
- Trello
- Slack
- Snoop Dogg
- Jake the Dog and Finn the Human
- The dog I once played with while walking home
- Chip (dog)
- Sonny and Wilson (dogs)
- Off Center
- Neal
- Neil deGrasse Tyson
- Neil Patrick Harris
- Neil Diamond
- Neil Young
- Neil Armstrong
- Neal again
- Thor

- Tom Kenny
- Kenny McCormick (RIP)
- Kenny Loggins
- The Danger Zone and the brave souls who built that highway
- Jake from State Farm
- The Mr. Game! team
- The Game Grumps
- Roosterteeth
- The T.V. Series, "Too Many Cooks"
- Richard Stallman
- Richard Stallman's Beard
- Steve Jobs
- The City of Hoboken
- Professor Duchamp
- Bill Nye the Science Guy
- Bill Gates (post Microsoft)
- Mom
- Professor Klapholz

EOF