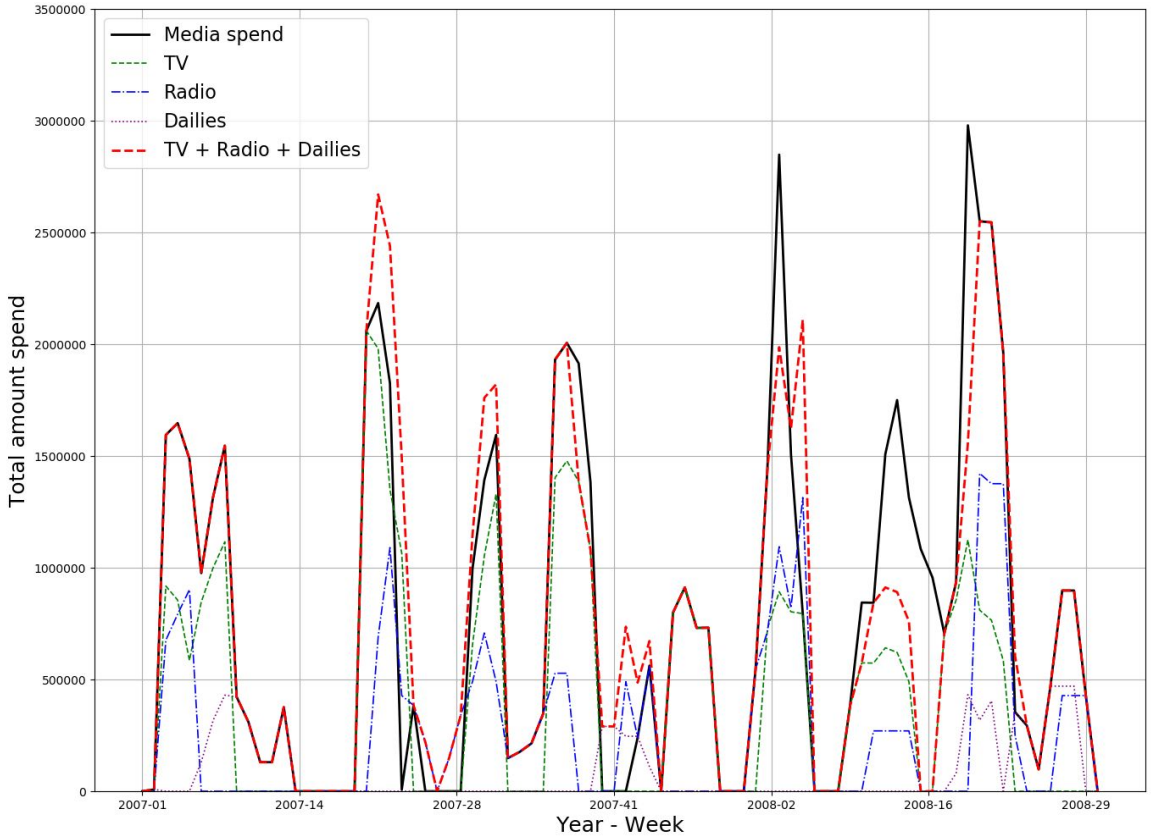
Regression explaining sales

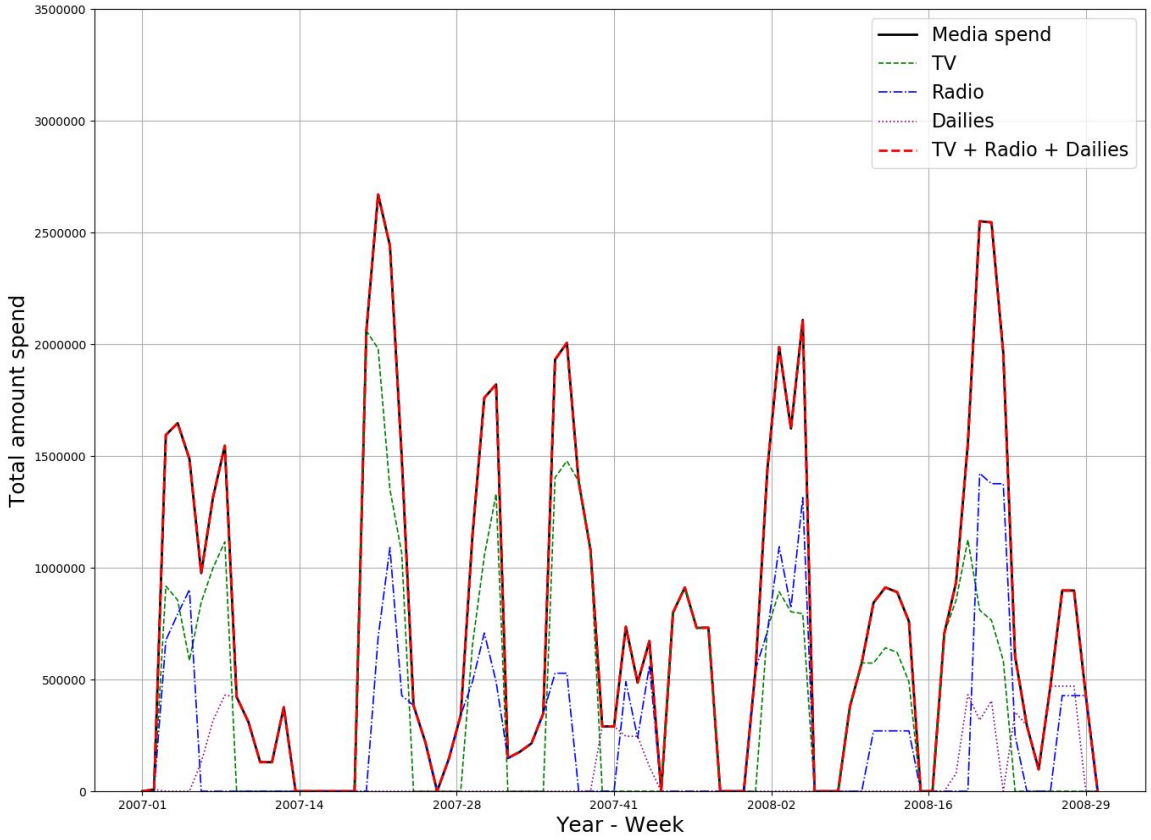
Mikkel Jensen

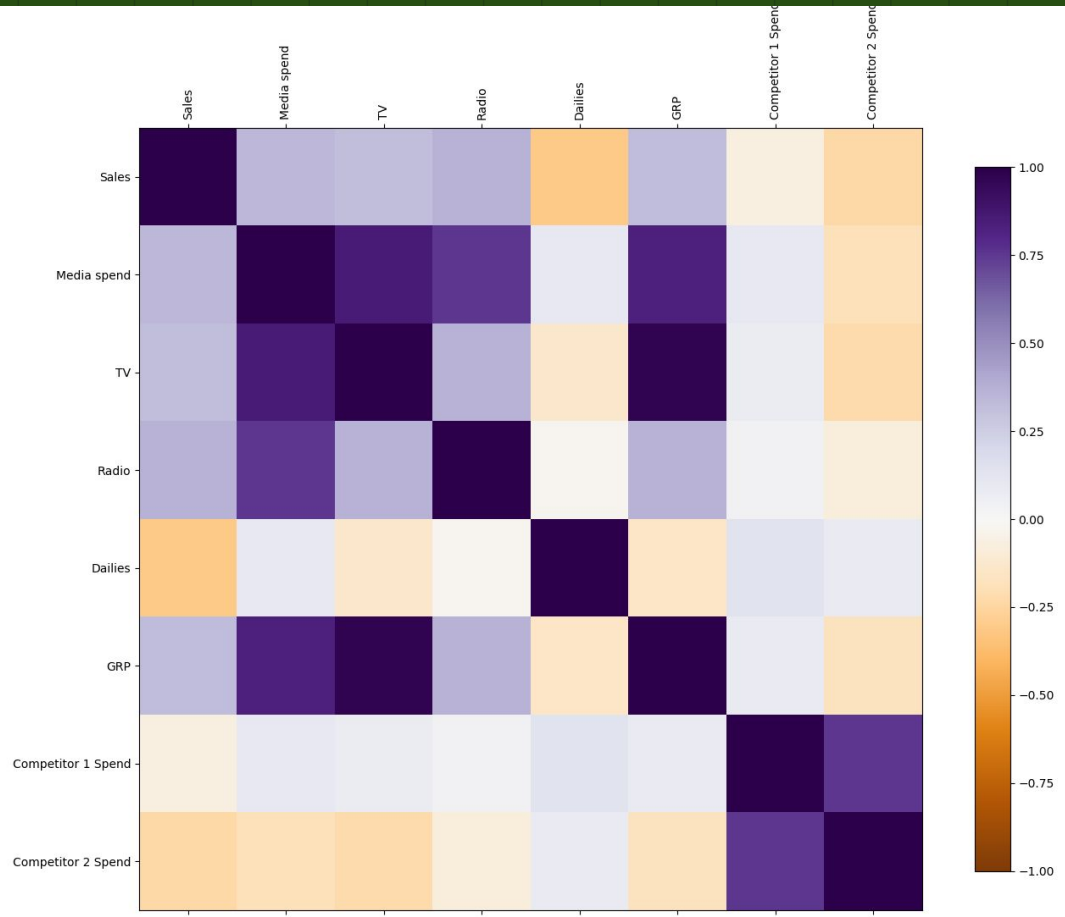
02 / 04 / 2019

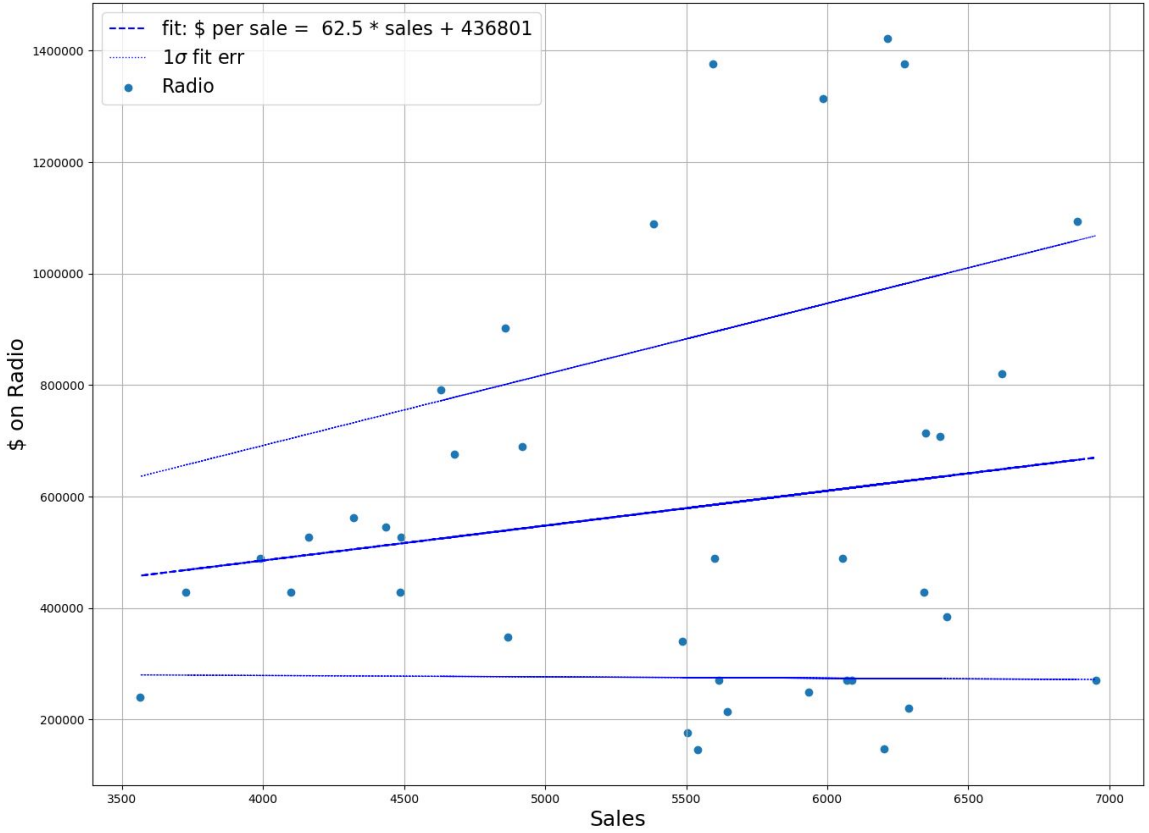

Media spend
TV
Radio
Dailies
TV + Radio + Dailies
Total amount spend
Year - Week
3500000
3000000
2500000
2000000
1500000
1000000
500000
0
2007-01
2007-14
2007-28
2007-41
2008-02
2008-16
2008-29

Data errors?

- TV + Radio + Dailies ≠ Media Spend
- Negative expenses
- Unrealistically small numbers (1 spend on TV)
- GRP > 0 when nothing is spend on TV


Media spend
TV
Radio
Dailies
TV + Radio + Dailies
Total amount spend
Year - Week
3500000
3000000
2500000
2000000
1500000
1000000
500000
0
2007-01
2007-14
2007-28
2007-41
2008-02
2008-16
2008-29


Sales
Media spend
TV
Radio
Dailies
GRP
Competitor 1 Spend
Competitor 2 Spend
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00


fit: $ per sale = 62.5 * sales + 436801
1σ fit err
Radio
$ on Radio
Sales
1400000
1200000
1000000
800000
600000
400000
200000
3500
4000
4500
5000
5500
6000
6500
7000

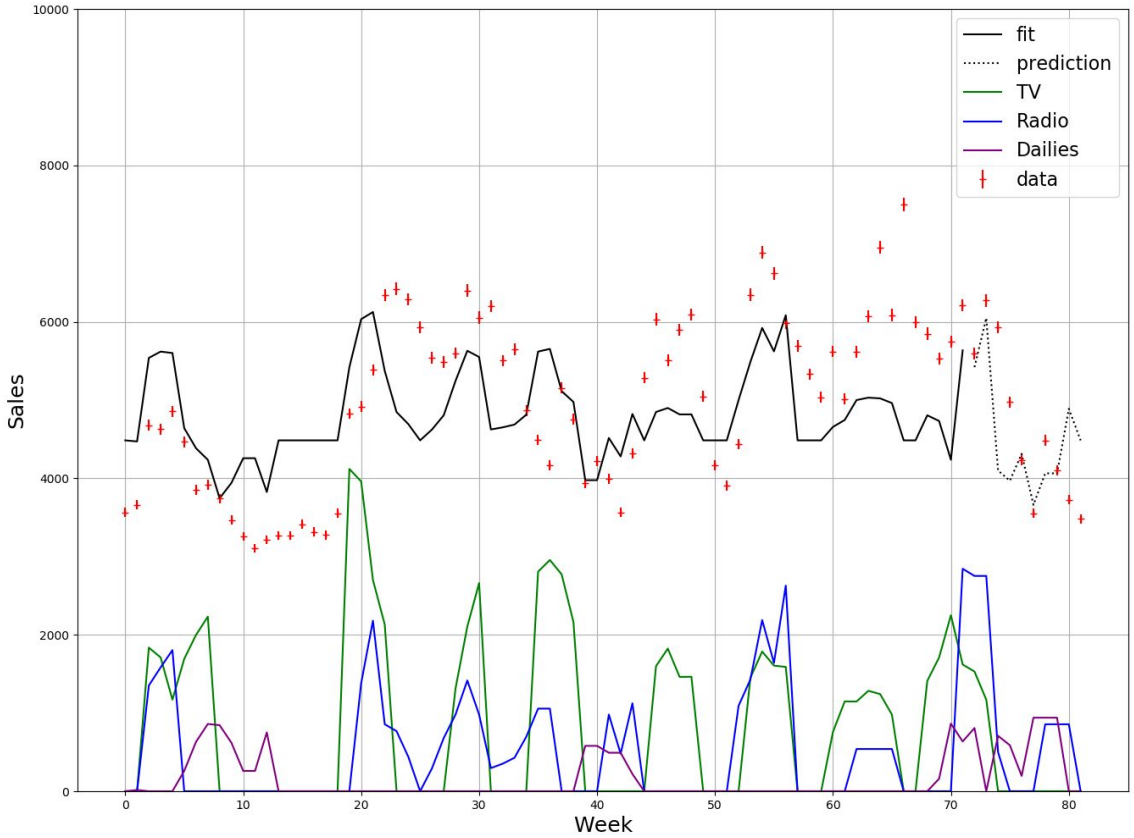
02 / 04 / 2019

Best fit: (ax + by + cz + d)

Sales no spend = 4485

	$/Sale
TV	2210
Radio	1060
Daily	-570

fit reduced chi2 = 201
Test data chi2 = 1480



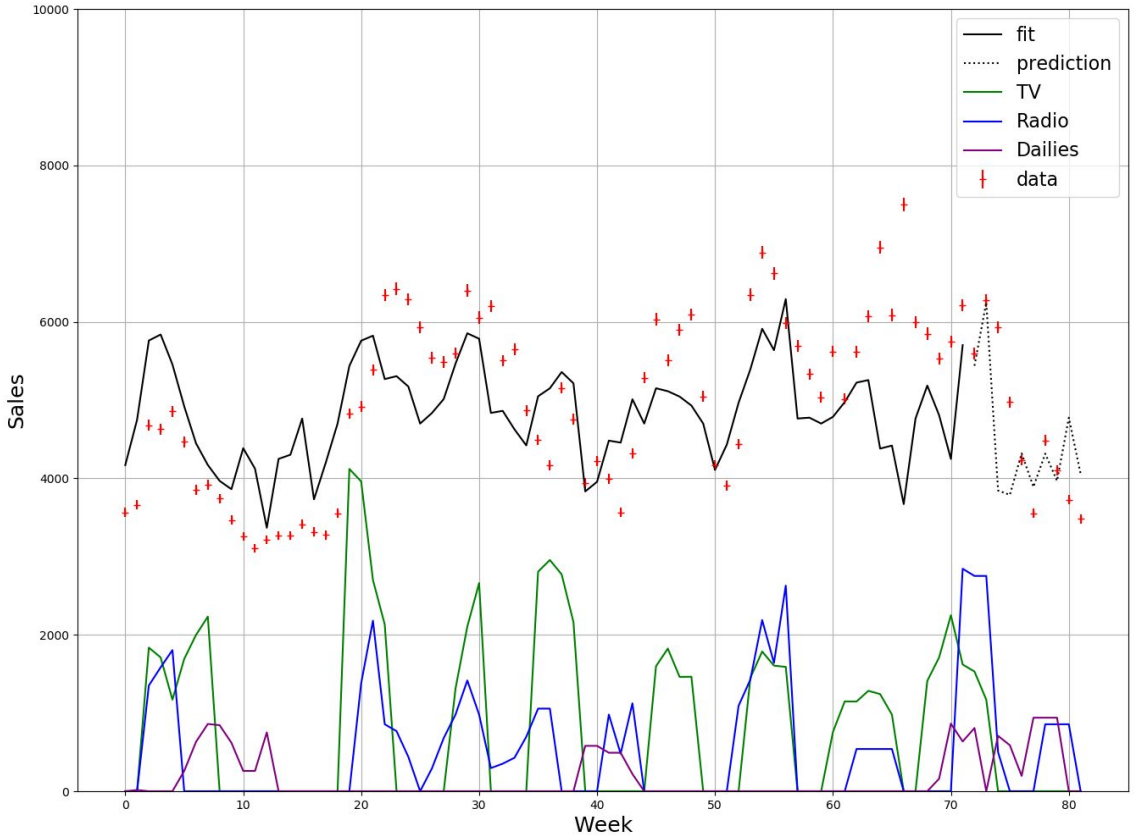
02 / 04 / 2019

Best fit: (hyperplane with Competitor spend)

Sales no spend = 4700

	$/Sale
TV	4410
Radio	1080
Daily	-530

fit reduced chi2 = 188
Test data chi2 = 1458



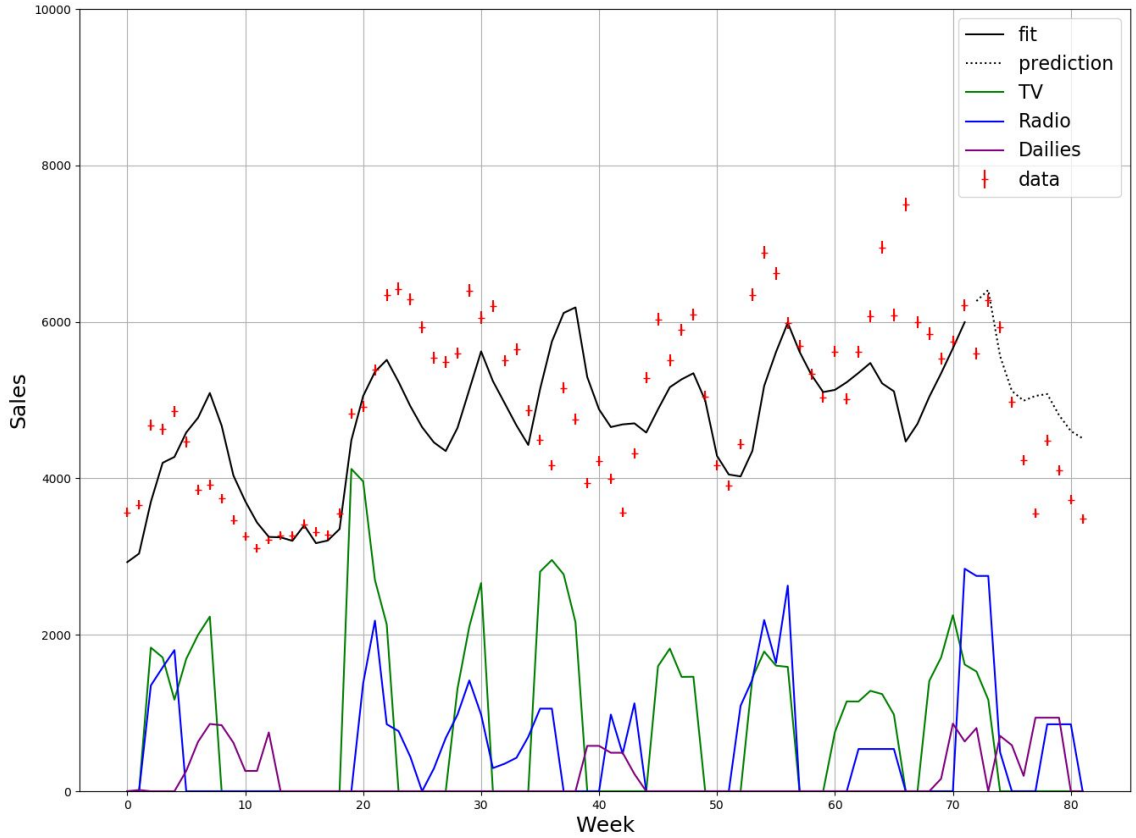
02 / 04 / 2019

Best fit: $\Sigma_{var}(M_{var,i} + B_{var,i} * A_{var,i})$

Sales no spend = 3117

	B	$/Sale
TV	0.73	1830
Radio	0.97	4900
Daily	0	0

fit reduced chi2 = 168
Test data chi2 = 1590



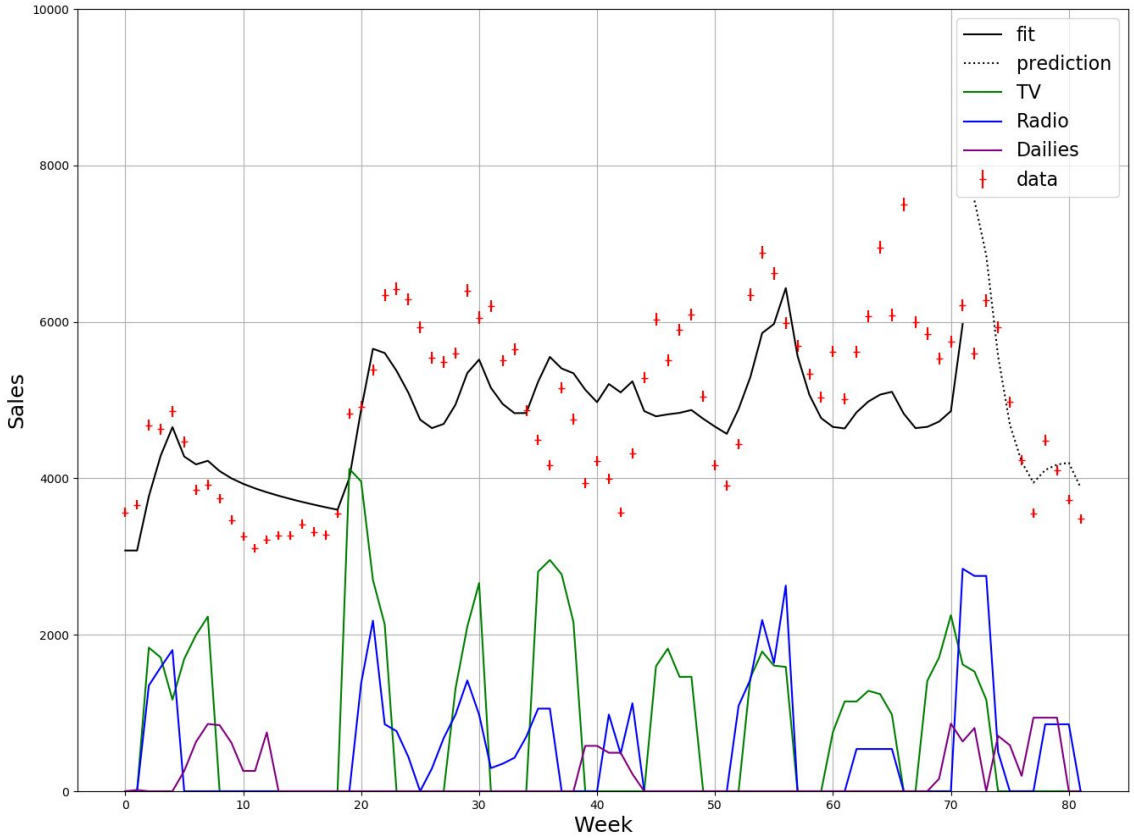
02 / 04 / 2019

Best fit: $\Sigma_{var}(M_{var,i} + B_{var,i} * A_{var,i})$

Sales no spend = 3077

	B	$/Sale
TV	0.94	4728
Radio	0.53	1353
Daily	0	0

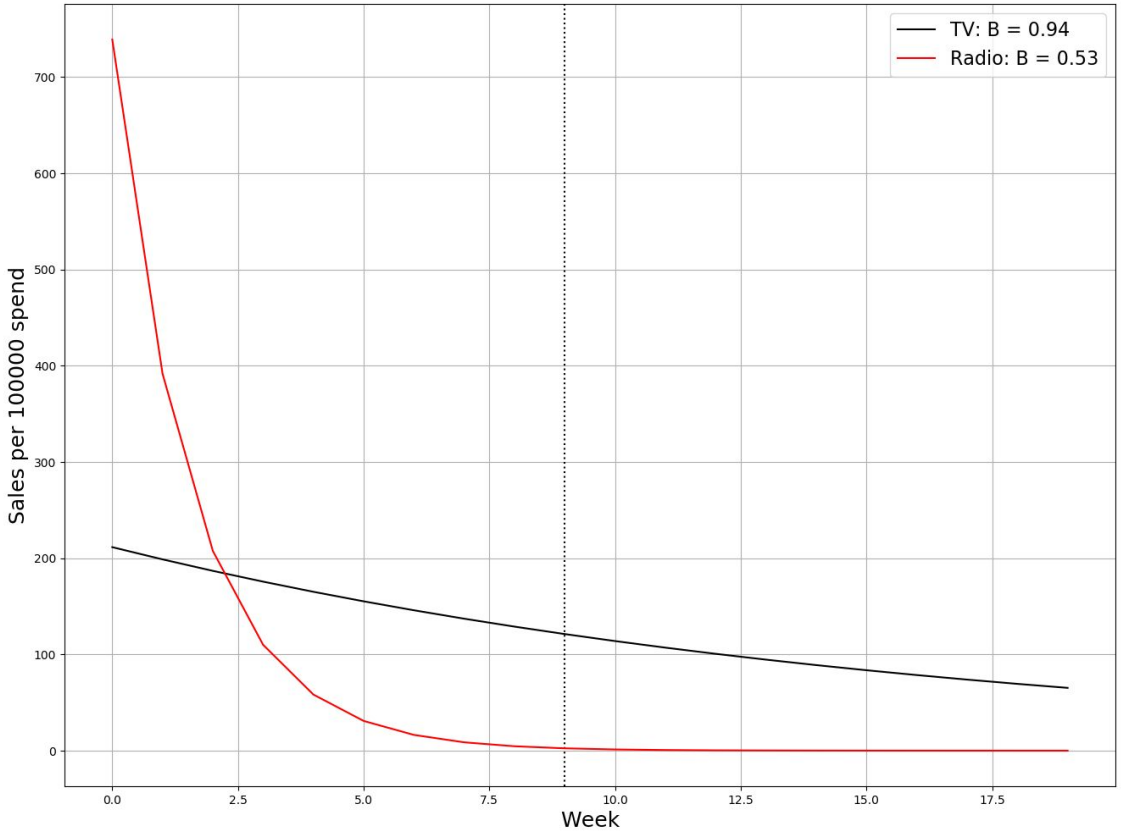
fit reduced chi2 = 151
Test data chi2 = 958



02 / 04 / 2019

Most efficient channel:
Radio if less than 9 weeks left, otherwise TV

	B	$/Sale
TV	0.94	4728
Radio	0.53	1353
Daily	0	0



Analysis results

- ~34% of sales are driven by media
- Radio is the most effective short term channel (up to nine weeks)
- Dailies seem to be a waste of money, but might be hard to fit because of the small amount invested

Suggested improvements

- Clean data more e.g the highest sales look wrong
- Fit sales not governed by commercials with a polynomial
- Play around with fit more, including all variables should be able to fit better
- Use other techniques as proper error estimation, machine learning, use priors while fitting campaigns separately

Hyperplane

+ 4061 C1S
- 1806 C2S

Hyperpoly

- 36607 CS1, B = 0.4
- 6065 CS2, B = 0.38