Glmnet for finance text regression

From ARKwiki

by brendan o'connor, december 2009 or so

See also GLMNet Runner instructions

Contents

Report

[10701 Report on Lasso, NMF, and LDA (http://www.ark.cs.cmu.edu/ARKwiki/images/f/f4/Lasso_nmf_lda_text_regression.pdf)] -- has a nice shorter form of the below (plus other stuff)

Overview

Using the glmnet R package

- CRAN (http://cran.r-project.org/web/packages/glmnet/index.html)
- Friedman, Hastie, Tibshirani april 2009 tech report (http://www-stat.stanford.edu/~hastie/Papers/glmnet.pdf)
- see also ESL Chapter 3 (http://www-stat.stanford.edu/~tibs/ElemStatLearn/) , and any of Trevor Hastie or Rob Tibshirani's recent presentations .. e.g. Tibshirani's 11/09 Google Pittsburgh presentation, or Hastie's useR keynote

Actually using this as a straight-up lasso model: least-squares linear regression with L1 regularization penalty on the weights. (So not using the elastic net penalty.)

The nice thing about glmnet is its tracing out of the entire warm-start regularization path, and that it's quite fast. The bad thing is that it crashes ("fortran memory error") with the fully sized datasets.

In all cases, response variable is stock log-volatility for future 12 months.

Train/Dev/Test splits

Using the same basic structure as the 2008 NAACL paper. For each of 2001-2006 as target YEAR,

- Dev mode:
  - Train on YEAR-5 .. YEAR-2 and predict YEAR-1 as dev set.
  - Tune lambda through its entire path here, looking at dev set error.
- Test mode:
  - Train on YEAR-5 .. YEAR-1 and predict YEAR as test set, using best lambda from before.
  - Report this MSE

Here is an equation-diagram. `F()` is the L1-penalized training MSE objective the linear regression package optimizes. P1,P2 are performance numbers of held-out MSE. M1 and M2 are particular models.

```
[P1] |---------- final performance in the forecasting setting: no tuning allowed ----------|

[M1]         |-------------------------- final model training ---------------------------|

                                    |---------------- lambda tuning --------------------|

[P2]                                                |------- dev holdout performance ----|

[M2]                                                        |---- dev set training ----|

     MSE[y6, argmax_w' F(w' | y1-5, argmin_lambda MSE[y5, argmax_w F(w | y1-4, lambda) ] ) ]
```

There are interesting connections to Kalman Filters? Reinforcement learning?

Data munging

glmnet needs data to be in a funny triples format. Making the featurename=>featureid mapping is annoying because it has to hold across year splits.

Everything is based on the publicly released version of the 10K dataset here (http://www.ark.cs.cmu.edu/10K/) . I had to clean up the volatility data a bit (there were duplicates, harmless for certain programs but deadly to R), and also had to reimplement the tokenizer, in a fun little data/code adventure. The plus side is this supports a cool highlighting-enabled model inspector UI (not installed anywhere, sorry).

Feature count thresholding

For unigrams, the full data size is 220k types over 30k examples at 0.5% non-sparsity. If you load all features, some of the splits make glmnet crash. It's not purely a memory issue -- under certain circumstances it can work, so the behavior feels bug-like. I exchanged a few emails with Hastie and Friedman, who were able to solve part of my problem, but not all of it. It is nice of them that they helped me.

I spent some time staring at Friedman's crazy Fortran code (see my blog post (http://anyall.org/blog/2009/07/fft-friedman-fortran-tricks/) from last summer), but found it was easier to do simple count thresholding.

At fc>=5 applied globally to all splits, number of features is more like 30k-50k. Performance seems similar as laxer thresholds fc>=2 or so.

The bigram features use a different strategy, because the feature space is really large and problematic (5.8 million raw). Every split chooses its own threshold to end up with 30k features. The thresholds for this are, for the 5-year-windows before 2001-2006: 162 193 236 308 380 466. Yes, this seems very high! I bet good features might be lurking under these thresholds.

This is one reason I wanted to do a variable ranking by F-test (the correct analogue of t-test or chi-square ranking for this setting, I believe) but haven't gotten around to implementing yet. It may, of course, be better to just write our own lasso regression solver -- it really shouldn't be that hard.

Penalties

L1 regularization for all text features.

The non-text feature, previous 12 months' volatility, is *not* penalized. This doesn't make much of a difference to prediction accuracy, but might be nice to use when lots more arbitrary non-text features get introduced.

Results: Performance

- glmnet bigrams+history does best
- run26 vs naacl08_log1p is a fair comparison: glmnet looks same or better
- glmnet does worse in text-only mode "nohist" ... run29_nohist vs. naacl08_nohist_log1p is the fair comparison.

	2001	2002	2003	2004	2005	2006	wtd avg	
run33_bigrams_30k	0.1801	0.1512	0.1569	0.1171	0.1209	0.1283	0.1406	log1p bigrams

run38_tf_bg	0.1854	0.1546	0.1630	0.1218	0.1222	0.1268	0.1437	tf bigrams
run26_nt_no_penalty	0.1849	0.1542	0.1713	0.1207	0.1271	0.1360	0.1474	log1p unigrams
run37_tf_ug	0.1926	0.1602	0.1874	0.1259	0.1277	0.1373	0.1536	tf unigrams
naacl08_log1p_bigrams	0.1852	0.1792	0.1599	0.1352	0.1307	0.1448	0.1538	
naacl08_tf	0.1885	0.1616	0.1925	0.1230	0.1272	0.1402	0.1541	
naacl08_log1p	0.1846	0.1764	0.1671	0.1309	0.1319	0.1458	0.1542	
guesshist	0.1747	0.1600	0.1873	0.1442	0.1365	0.1463	0.1576	
ols_justhist	0.1922	0.1631	0.2040	0.1272	0.1275	0.1403	0.1578	
svr_justhist_0bias	0.2053	0.1653	0.2051	0.1337	0.1405	0.1517	0.1655	
ols_justhist_0bias	0.2056	0.1655	0.2073	0.1344	0.1392	0.1520	0.1659	
naacl08_nohist_log1p_bigrams	0.1968	0.2015	0.1729	0.1500	0.1394	0.1532	0.1667	
naacl08_nohist_log1p	0.2107	0.2214	0.2040	0.1693	0.1581	0.1715	0.1873	
run29_nohist	0.2142	0.2433	0.2339	0.2033	0.1764	0.1817	0.2078	log1p unigrams, no history
variance	0.4106	0.4238	0.3539	0.2906	0.2773	0.2893	0.3354	
svr_justhist_wtf	0.2433	0.4323	0.1869	0.2717	0.3184	5.6778	1.2058	

Results: Sparseness

The glmnet models are pretty sparse. Every train split starts off with about 30k features, after count thresholding.

- Pure-text model uses the most features.
- Bigram models use more features than the unigram model.

```
**********  run29_nohist  **********
  year final_mse final_size dev_best_size chosen_lambda
1 2001 0.2142127       1936          1967   0.005706636
2 2002 0.2433141        622           708   0.010803411
3 2003 0.2339275       2088          2075   0.005773097
4 2004 0.2033474       1271          1402   0.007746903
5 2005 0.1763593       1598          1713   0.006742138
6 2006 0.1816579       1314          1431   0.007362409
**********  run26_nt_no_penalty  **********
  year final_mse final_size dev_best_size chosen_lambda
1 2001 0.1848553        160           209    0.01194129
2 2002 0.1541742        152           185    0.01246913
3 2003 0.1713024        148           222    0.01225857
4 2004 0.1207476        218           280    0.01056943
5 2005 0.1271308         15            26    0.02836653
6 2006 0.1359893         33            63    0.01790349
```

```
**********  run33_bigrams_30k  **********
  year final_mse final_size dev_best_size chosen_lambda
1 2001 0.1801198        282           337   0.009903041
2 2002 0.1512366        232           279   0.010709062
3 2003 0.1568740        112           166   0.012736115
4 2004 0.1170910        310           411   0.009004996
5 2005 0.1209469        116           169   0.012385845
6 2006 0.1283334        111           168   0.010361138
```

Final models

- train_df: number of training documents this feature appears in
- test_df: number of testing documents this feature appears in
- the NT_before feature is the previous 12 months' log volatility.

Unigram and bigram models below. I love the bigrams. Much more interpretable, e.g. "merger_agreement". I think this bodes well for strategies of extracting syntactically-aware features (subtree spans (aka coherent n-grams) and dep paths)

bigrams (run33)

```
******************************************************
*                       2001                         *
******************************************************
final_test       2001
alpha    1
dfmax    5000
year_window      5
refine   TRUE
input_dir        /usr2/brendano/10K/feat_bigrams/prune_30k/2001.relprune
input_ext        tf.num
more_nontext_data_file
nohist   FALSE
outfile          results.RData
final_test       2001
final_train      1996 1997 1998 1999 2000
dev_test         2000
dev_train        1996 1997 1998 1999
                        name          beta train_df test_df
1            merger_agreement -1.262075e-01      530     123
2                    and_plan -6.777867e-02      360     111
3                   merger_is -4.826482e-02      245      45
4               these_declines -4.550718e-02      190      38
5                 tender_offer -2.955451e-02      444     115
6                 equity_ratio -2.944191e-02      218      38
7               mortgage_notes -2.892167e-02      237      49
8                not_represent -2.720643e-02      433     138
9                   and_margin -2.694519e-02      203      52
10                tax_planning -2.490926e-02      254      79
11              for_stockbased -2.336148e-02      695      25
12                     to_vote -2.307104e-02      162      54
13                  mature_or -2.174772e-02      186      63
14                       reit -2.067455e-02      401     107
15          after_adjustments -2.066443e-02      178      40
16             with_generally -2.040548e-02      920     212
17                  of_merger -1.842385e-02      403     107
18        shortterm_borrowing -1.798213e-02      253      46
```

19	nearly	-1.796475e-02	917	217	
20	information_required	-1.571677e-02	226	71	
21	rental_rates	-1.565656e-02	296	68	
22	interest_increased	-1.546737e-02	164	45	
23	rates_are	-1.518478e-02	953	330	
24	awarded	-1.504190e-02	694	164	
25	recall	-1.474719e-02	197	61	
26	rising_interest	-1.467110e-02	476	178	
27	weather	-1.428674e-02	1644	372	
28	than_#%	-1.393158e-02	2111	574	
29	positive	-1.359695e-02	2612	660	
30	lifo	-1.308082e-02	256	49	
31	per_unit	-1.210631e-02	828	192	
32	has_sufficient	-1.111098e-02	375	74	
33	financing_needs	-1.109461e-02	274	47	
34	from_property	-1.084161e-02	165	33	
35	warmer	-1.067608e-02	186	37	
36	association_of	-1.047833e-02	509	122	
37	property_management	-1.042916e-02	273	53	
38	the_leasing	-1.035389e-02	197	52	
39	general_partner	-1.018262e-02	381	75	
40	properties	-1.000893e-02	2816	668	
41	bulk	-9.157148e-03	590	137	
42	net_income	-8.988949e-03	7817	1769	
43	about_$#	-8.928137e-03	310	56	
44	improved	-8.908302e-03	5157	1093	
45	#_accounting	-8.773142e-03	3966	1522	
46	board_of	-8.378064e-03	4086	1118	
47	nm	-7.525327e-03	172	39	
48	reflected_a	-7.016303e-03	540	130	
49	favor	-7.001535e-03	619	165	
50	the_pending	-6.962767e-03	287	68	
51	measure	-6.905550e-03	2109	683	
52	its_shortterm	-6.795933e-03	424	113	
53	eligibility	-6.745637e-03	167	58	
54	partnership_s	-6.605457e-03	238	47	
55	distributions	-6.465895e-03	1621	375	
56	#%_reduction	-6.412121e-03	384	74	
57	unsecured	-6.323509e-03	2518	531	
58	adopted	-6.201854e-03	3535	1155	
59	s_policy	-5.790030e-03	867	196	
60	income_from	-5.572659e-03	4198	1016	
61	business_risk	-5.532348e-03	253	50	
62	repayments	-5.520166e-03	1728	506	
63	a_return	-5.454681e-03	324	78	
64	by_operations	-5.268525e-03	2070	455	
65	company_recently	-5.099339e-03	276	26	
66	million_lower	-5.098429e-03	256	73	
67	industrial	-5.034362e-03	2159	485	
68	or_less	-4.950085e-03	1254	418	
69	registrant	-4.896110e-03	806	159	
70	pretax_earnings	-4.688069e-03	486	118	
71	to_apply	-4.558966e-03	495	127	
72	net_earnings	-4.052990e-03	1425	281	
73	the_transactions	-4.038925e-03	400	140	
74	share_in	-4.012855e-03	2702	604	
75	and_contingency	-3.961676e-03	420	4	
76	operating_investing	-3.792223e-03	219	59	
77	rate	-3.721644e-03	10025	2495	
78	in_real	-3.632273e-03	518	147	
79	increases	-3.628979e-03	9072	2194	
80	partnership_the	-3.624495e-03	296	57	
81	utility	-3.541967e-03	1051	197	
82	higher_in	-3.536753e-03	967	229	
83	longterm	-3.510588e-03	6899	1750	
84	dividend_payments	-3.319656e-03	610	143	
85	gallons	-3.148707e-03	179	32	

86	requirements_of	-3.091578e-03	1940	512	
87	per_share	-3.085432e-03	7205	1690	
88	taxexempt	-2.915238e-03	692	168	
89	income	-2.863488e-03	10921	2571	
90	obtained_in	-2.570344e-03	429	100	
91	the_seller	-2.482410e-03	264	72	
92	local_market	-2.326142e-03	197	63	
93	company_periodically	-2.267623e-03	301	72	
94	real	-2.186836e-03	2810	730	
95	agreement_provides	-2.089532e-03	802	167	
96	computed_in	-1.992677e-03	230	47	
97	year_s	-1.902076e-03	1378	279	
98	third_quarters	-1.803058e-03	678	158	
99	land_for	-1.713351e-03	168	43	
100	#_dividends	-1.696760e-03	449	108	
101	of_food	-1.609705e-03	272	58	
102	ffo	-1.443037e-03	234	65	
103	normal	-1.414176e-03	4204	864	
104	offer_the	-1.404257e-03	303	89	
105	testing_is	-1.350799e-03	186	13	
106	increased_in	-1.264886e-03	2688	572	
107	for_longlived	-1.260973e-03	712	129	
108	acquire_all	-1.245188e-03	171	47	
109	s_securities	-1.200628e-03	465	118	
110	estate_taxes	-1.058092e-03	333	86	
111	fixedrate	-1.051551e-03	594	259	
112	funds_for	-8.677312e-04	1182	280	
113	environmental_matters	-3.970190e-04	692	126	
114	approval	-3.922519e-04	2986	832	
115	paid_on	-3.225214e-04	1396	421	
116	and_property	-1.174492e-04	712	159	
117	adopted_the	-8.470290e-05	704	247	
118	of_gas	-3.119037e-05	301	68	
119	the_independent	-2.652579e-05	298	66	
120	exploring	2.232555e-05	413	123	
121	company_has	4.560720e-05	8965	1531	
122	was_determined	5.069976e-05	556	254	
123	structure_of	1.246850e-04	481	121	
124	security_agreement	1.434046e-04	292	83	
125	its_inventory	1.665742e-04	460	67	
126	the_sum	2.239150e-04	761	170	
127	versus_the	3.083442e-04	624	169	
128	fixed_assets	3.244169e-04	1331	388	
129	covenants_in	3.821478e-04	381	118	
130	pay_the	4.572334e-04	948	248	
131	computer_systems	4.686935e-04	3134	175	
132	of_approximately	5.283194e-04	8078	2026	
133	goodwill	5.345267e-04	3202	1203	
134	will_be	5.414537e-04	10169	2393	
135	dispute	5.912974e-04	520	161	
136	companies_acquired	6.375294e-04	241	66	
137	general_and	6.444954e-04	8462	2079	
138	not_currently	6.538495e-04	1707	483	
139	the_expenses	7.087923e-04	770	201	
140	determined_that	7.146917e-04	1832	484	
141	material	7.709210e-04	9476	2211	
142	to_support	8.784625e-04	4629	1264	
143	be_material	8.848804e-04	2135	290	
144	and_operational	9.032991e-04	841	209	
145	the_year	9.854292e-04	9826	2235	
146	amendment_to	1.004549e-03	552	208	
147	establishment_of	1.210717e-03	1155	302	
148	computer	1.246894e-03	6525	927	
149	lower_margins	1.275976e-03	908	218	
150	that_such	1.277436e-03	3522	693	
151	sold_as	1.294290e-03	593	126	
152	for_related	1.377425e-03	365	18	

153	insufficient_to	1.425314e-03	396	155
154	overhead	1.480042e-03	2768	698
155	expenses	1.525199e-03	10917	2562
156	placement_of	1.549921e-03	1156	285
157	subsidiaries_were	1.640142e-03	179	42
158	auto	1.731347e-03	380	89
159	customers	1.787397e-03	7538	1813
160	personnel	1.832473e-03	5746	1495
161	underperforming	1.896756e-03	350	97
162	expenses_incurred	1.921094e-03	1675	420
163	efforts	2.222835e-03	6745	1514
164	our_revenues	2.312892e-03	462	555
165	is_attributed	2.362639e-03	405	94
166	expect	2.376394e-03	4148	1481
167	guaranteed_by	2.379324e-03	892	209
168	sell_or	2.390994e-03	314	114
169	and_focus	2.663327e-03	199	56
170	to_goodwill	2.736819e-03	371	224
171	product_offerings	2.738033e-03	1149	373
172	funded_its	2.756988e-03	501	74
173	private	2.869530e-03	5564	1516
174	implementation_and	2.912171e-03	431	113
175	to_ensure	2.935285e-03	2586	402
176	for_operating	2.944100e-03	766	200
177	provider	3.040210e-03	1630	614
178	obligations_including	3.121923e-03	169	50
179	were_attributable	3.138907e-03	545	125
180	million_incurred	3.233818e-03	240	49
181	eligible	3.276076e-03	1553	403
182	not	3.472458e-03	10950	2586
183	cash_balance	3.686979e-03	479	153
184	months_during	3.767147e-03	167	43
185	company_is	3.798302e-03	8027	1391
186	acquisitions_that	3.802221e-03	305	123
187	of_salaries	3.802453e-03	549	266
188	accounts_receivables	4.235059e-03	272	89
189	additional_capital	4.326711e-03	1772	532
190	security_interest	4.327667e-03	440	106
191	stage	4.333026e-03	1278	379
192	fixtures	4.346162e-03	707	195
193	startup	4.397017e-03	2121	477
194	asset_impairment	4.477693e-03	335	182
195	targets	4.489880e-03	642	239
196	development_of	4.498658e-03	4775	1237
197	other_data	4.505275e-03	281	88
198	accounts_receivable	4.551909e-03	5218	1294
199	forwardlooking_statements	4.554508e-03	5876	1717
200	s_statement	4.576480e-03	297	59
201	operating_losses	4.711084e-03	2513	793
202	materially_from	4.716979e-03	5261	1445
203	inventory_in	4.841572e-03	541	127
204	expertise	4.955277e-03	993	321
205	and_services	4.989054e-03	3687	991
206	deficit_of	5.018769e-03	697	280
207	margin_the	5.025868e-03	524	116
208	received_proceeds	5.081413e-03	273	90
209	software	5.152839e-03	6269	1048
210	to_raise	5.158166e-03	1482	588
211	of_goodwill	5.198622e-03	1663	752
212	capital_or	5.274886e-03	335	103
213	in_equal	5.343039e-03	275	63
214	software_which	5.370027e-03	385	45
215	not_in	5.404752e-03	1004	286
216	travel_expenses	5.532610e-03	351	119
217	department_and	5.580229e-03	179	45
218	date	5.678281e-03	8490	2090
219	overhead_and	5.678965e-03	436	124

220	consulting	5.859536e-03	2702	783
221	warrants	5.968116e-03	2209	664
222	assurances_can	6.006718e-03	492	80
223	infrastructure	6.018253e-03	2477	801
224	be_successful	6.289043e-03	1847	545
225	amendment	6.320686e-03	1270	515
226	the_excess	6.393019e-03	1071	310
227	debenture	6.457115e-03	348	77
228	on_form	6.585171e-03	2197	659
229	and_iii	6.692083e-03	1885	319
230	financing_to	6.792101e-03	930	271
231	year_#	6.987510e-03	8266	1371
232	we_began	7.411176e-03	270	301
233	duplicative	7.430653e-03	172	42
234	cash_to	7.556147e-03	1195	350
235	gold	7.579309e-03	246	37
236	obtain_additional	7.909861e-03	955	334
237	additional	7.965007e-03	10346	2461
238	we_were	8.323871e-03	448	549
239	covenants	9.213412e-03	3965	1004
240	earlier_periods	9.294714e-03	186	14
241	operations_during	9.805673e-03	1847	434
242	discontinuation_of	9.907559e-03	321	92
243	purchasing	1.044331e-02	2510	654
244	focused_on	1.048374e-02	1346	396
245	with_sufficient	1.050338e-02	187	50
246	default_under	1.076907e-02	415	129
247	loss_for	1.105741e-02	1722	496
248	major_vendors	1.119090e-02	171	5
249	noncash_charge	1.137295e-02	631	200
250	and_b	1.138662e-02	548	125
251	company_experienced	1.161394e-02	1581	227
252	a_writedown	1.184081e-02	328	118
253	significant_losses	1.233789e-02	274	127
254	related_parties	1.266013e-02	263	63
255	a_waiver	1.301551e-02	392	110
256	financing_or	1.307028e-02	499	174
257	china	1.390182e-02	586	180
258	additional_financing	1.394465e-02	1643	573
259	negotiating	1.399636e-02	831	209
260	in_accounts	1.402673e-02	3159	805
261	securitizations	1.407101e-02	176	124
262	accounting_in	1.431780e-02	197	73
263	convertible	1.453411e-02	2869	777
264	company_recognizes	1.508724e-02	647	103
265	whether_a	1.635947e-02	249	231
266	administrative	1.797145e-02	9765	2310
267	company_accounted	1.826889e-02	173	29
268	net_loss	1.938623e-02	4124	1205
269	raising	1.984908e-02	653	230
270	negative_cash	1.987066e-02	427	177
271	lender_the	2.045615e-02	219	46
272	for_services	2.243303e-02	738	196
273	recruiting	2.273032e-02	763	321
274	raise_capital	2.417492e-02	211	124
275	defaults	2.609486e-02	454	133
276	reporting_comprehensive	3.347841e-02	1302	7
277	business_a	3.401208e-02	198	55
278	financial_covenants	3.466648e-02	1604	457
279	risk_not	3.970263e-02	772	74
280	its_computer	5.200126e-02	1322	35
281	a_going	1.009885e-01	295	115
282	NT_before	7.792664e-01	11075	2596

Bias term:

s0
-0.7260307

****

```
*                      2002                                *
*********************************************************
final_test       2002
alpha    1
dfmax    5000
year_window      5
refine   TRUE
input_dir        /usr2/brendano/10K/feat_bigrams/prune_30k/2002.relprune
input_ext        tf.num
more_nontext_data_file
nohist   FALSE
outfile          results.RData
final_test       2002
final_train      1997 1998 1999 2000 2001
dev_test         2001
dev_train        1997 1998 1999 2000
                      name          beta train_df test_df
1          merger_agreement -9.128067e-02      603     131
2                merger_is -7.320750e-02      260      46
3                 and_plan -7.309385e-02      435     158
4         after_adjustments -5.025023e-02      207      45
5             not_represent -4.223465e-02      547     152
6             million_lower -3.108978e-02      291     118
7       information_required -2.505463e-02      281      75
8                of_merger -2.296292e-02      459     105
9                  weather -2.033912e-02     1814     438
10       cash_distributions -1.981122e-02      326      97
11             computed_in -1.830218e-02      265      60
12      property_management -1.746377e-02      306      62
13                 supreme -1.685129e-02      237      66
14                    reit -1.582573e-02      487     105
15            #_accounting -1.552960e-02     4812    2145
16        accumulated_other -1.507139e-02      363     297
17             estate_taxes -1.503135e-02      388     117
18             rental_rates -1.446598e-02      339     100
19                   recall -1.401818e-02      247      96
20                     rate -1.285521e-02    11310    2759
21           mortgage_notes -1.202954e-02      255      55
22               net_income -1.192275e-02     8609    1997
23         which_represents -1.125678e-02     1000     316
24                   rising -1.095366e-02     1212     355
25          financing_needs -1.047882e-02      292      83
26                made_from -1.032297e-02      446     130
27      operating_investing -9.112643e-03      252      69
28                repayments -9.058190e-03     2047     617
29               repurchases -9.042250e-03     1248     466
30           authorized_the -8.469541e-03      880     274
31               properties -8.284737e-03     3117     842
32                   tender -7.944975e-03      915     271
33            distributions -7.728196e-03     1857     458
34              consumption -7.691204e-03      406     164
35           association_of -7.587238e-03      575     119
36             tender_offer -7.414754e-03      534     173
37           unsecured_line -7.399506e-03      324      48
38                    rates -7.291336e-03    10202    2734
39                #_goodwill -7.072936e-03      573    1949
40                   than_#% -6.785414e-03     2442     822
41                 units_and -6.470081e-03      783     279
42           from_insurance -6.313474e-03      211      59
43             about_market -6.256087e-03     7219    2568
44           were_generally -6.129905e-03      262      72
45            its_shortterm -6.031854e-03      499     108
46                   year_s -5.906347e-03     1442     387
47    accounting_principles -5.898531e-03     2591    1935
48                  #_weeks -5.696095e-03      516     169
49               income_from -5.331844e-03     4718    1196
50           would_decrease -5.285039e-03      467     215
```

51	authorization	-4.866353e-03	784	272
52	planning_strategies	-4.855822e-03	193	295
53	increases	-4.789230e-03	10112	2411
54	generic	-4.748683e-03	194	59
55	board_of	-4.588089e-03	4767	1478
56	other_intangible	-4.567674e-03	871	1967
57	fixedrate	-4.543050e-03	812	303
58	improved	-4.339357e-03	5524	1315
59	gains	-4.300337e-03	5324	1548
60	market	-4.295358e-03	11512	2835
61	led_by	-4.190133e-03	429	131
62	income	-4.007652e-03	12109	2823
63	responsible_parties	-4.003529e-03	196	47
64	market_risk	-3.923513e-03	7824	2700
65	regulatory	-3.751120e-03	4633	1301
66	or_less	-3.567533e-03	1553	586
67	adopted	-3.348695e-03	4214	1676
68	net_earnings	-3.130686e-03	1494	341
69	rate_changes	-3.090466e-03	1597	582
70	this_pronouncement	-3.062730e-03	273	142
71	apartment	-3.029747e-03	229	47
72	the_equity	-2.842120e-03	1844	640
73	utility	-2.743609e-03	1136	312
74	per_unit	-2.672440e-03	919	268
75	excluding	-2.427645e-03	5722	1590
76	a_return	-2.417313e-03	357	134
77	rising_interest	-2.412374e-03	584	156
78	nearly	-2.318608e-03	1028	269
79	dividend	-2.250816e-03	3427	896
80	and_liability	-2.158991e-03	855	329
81	both	-2.086388e-03	9472	2429
82	share_in	-1.903946e-03	2947	734
83	#year	-1.894525e-03	980	303
84	unsecured	-1.854874e-03	2768	658
85	removed	-1.667657e-03	485	232
86	approvals	-1.576838e-03	1620	458
87	commodity	-1.516380e-03	1761	603
88	considers	-1.452238e-03	1736	639
89	an_unsecured	-1.450969e-03	665	125
90	favor	-1.398315e-03	716	266
91	#%_in	-1.367671e-03	8906	2077
92	cumulative_effect	-1.358364e-03	1071	771
93	useful	-1.351319e-03	1473	1545
94	reinvestment	-1.185861e-03	728	217
95	strong	-1.165798e-03	3484	916
96	by_higher	-9.276402e-04	1671	484
97	declared_a	-8.272377e-04	434	102
98	commodity_prices	-7.915855e-04	724	292
99	share	-7.452245e-04	9697	2432
100	had_entered	-7.343815e-04	330	98
101	changes_in	-6.846181e-04	9667	2661
102	#_$	-6.328193e-04	5193	1992
103	price_increases	-5.515247e-04	1463	379
104	approval_of	-5.410762e-04	1692	556
105	was_due	-5.371587e-04	6949	1797
106	growth_was	-4.774066e-04	1008	241
107	were_partially	-4.621441e-04	2823	818
108	the_recognition	-3.399417e-04	1782	802
109	dividends	-1.811378e-04	5536	1407
110	more_than	-1.528559e-04	4616	1355
111	rates_are	-1.440143e-04	1194	422
112	award	-9.757902e-05	806	237
113	in_market	-8.936418e-05	2459	1063
114	these_increases	-6.844263e-05	2637	647
115	of_eligible	2.629809e-05	648	165
116	experienced_any	3.022529e-05	899	133
117	loans_from	8.722495e-05	507	126

118	working_capital	1.527013e-04	9181	2210	
119	not_experienced	1.953034e-04	1140	213	
120	sell_or	2.385684e-04	406	173	
121	no_income	2.911395e-04	276	77	
122	which_had	2.967411e-04	1953	506	
123	office_in	3.465118e-04	556	189	
124	bandwidth	5.925200e-04	209	124	
125	long_distance	6.334117e-04	318	81	
126	but_there	6.515877e-04	448	118	
127	assigned_to	6.524635e-04	780	360	
128	acquired_substantially	6.584319e-04	353	79	
129	a_significant	7.635679e-04	7759	2146	
130	subordinated_notes	7.793242e-04	1386	374	
131	infrastructure	8.354580e-04	3156	972	
132	experienced	8.843292e-04	7046	1848	
133	operating_losses	1.010010e-03	3023	1070	
134	provided_the	1.057257e-03	852	229	
135	can_be	1.104623e-03	8923	2211	
136	issue	1.210616e-03	6091	1282	
137	costs_related	1.218604e-03	3864	1110	
138	startup	1.244959e-03	2389	454	
139	sales_net	1.289276e-03	1298	305	
140	acquisitions_as	1.298897e-03	661	161	
141	in_charges	1.349898e-03	259	112	
142	establishment_of	1.402843e-03	1350	389	
143	finance	1.741507e-03	6150	1497	
144	not_expected	1.924074e-03	2871	659	
145	company_has	2.109763e-03	9387	1651	
146	its_financial	2.117809e-03	2721	884	
147	ensure_that	2.130270e-03	1721	329	
148	there_can	2.295573e-03	6743	1423	
149	problems	2.327624e-03	4789	711	
150	its_operations	2.414216e-03	4499	722	
151	its_sales	2.658847e-03	1390	242	
152	additional_capital	2.676171e-03	2124	679	
153	corporate_overhead	2.747799e-03	515	137	
154	to_ensure	2.918330e-03	2859	589	
155	not_in	2.929804e-03	1198	375	
156	obtain_such	2.991683e-03	433	120	
157	personnel	3.042087e-03	6720	1737	
158	its_initial	3.084917e-03	1066	149	
159	the_company	3.157982e-03	11183	2394	
160	overhead	3.433257e-03	3120	909	
161	million_including	3.519640e-03	1572	490	
162	accounts_receivable	3.628091e-03	5852	1731	
163	a_foreign	3.641215e-03	818	222	
164	financing_to	3.646917e-03	1119	354	
165	placement_of	3.716512e-03	1317	334	
166	operations_during	3.833403e-03	2061	541	
167	be_successful	3.903372e-03	2267	723	
168	such_financing	3.918953e-03	854	219	
169	million_incurred	3.978831e-03	264	64	
170	gold	3.984030e-03	241	55	
171	expertise	4.031228e-03	1227	452	
172	for_services	4.088903e-03	873	348	
173	loss_for	4.105456e-03	2005	695	
174	increased_its	4.338183e-03	1549	236	
175	lender_the	4.361915e-03	236	57	
176	cash_to	4.467375e-03	1421	462	
177	compliant	4.514864e-03	3895	52	
178	systems_and	4.605061e-03	5513	884	
179	obligations_including	4.614800e-03	197	113	
180	in_default	4.653155e-03	461	224	
181	financing	4.673125e-03	10132	2538	
182	software	4.934879e-03	7036	1179	
183	s_statement	4.978321e-03	321	125	
184	an_event	4.995627e-03	444	296	

```
185               debenture  5.132523e-03      374      91
186                basis_as  5.281615e-03     1060     371
187    additional_financing  5.316950e-03     2082     665
188                 will_be  5.592289e-03    11371    2726
189               may_arise  5.863932e-03      722     156
190            financing_of  6.098220e-03     1180     246
191            inventory_in  6.174259e-03      597     249
192       product_offerings  6.242241e-03     1421     475
193            contemplates  6.244681e-03      227      72
194        computer_systems  6.247211e-03     3261     160
195                  travel  6.331463e-03     2021     721
196           no_assurance  6.350154e-03     7134    1592
197              or_equity  6.517620e-03     1199     420
198               computer  6.711492e-03     7106    1037
199             securitized  7.092638e-03      283      78
200              additional  7.405173e-03    11566    2752
201         company_issued  7.620853e-03     1908     289
202              that_such  7.668374e-03     3941     884
203          guaranteed_by  7.806233e-03     1002     289
204            loss_before  7.986041e-03     2408     725
205                  and_b  8.103410e-03      621     220
206       obtain_additional  9.193112e-03     1218     451
207             negotiating  9.542157e-03      927     261
208                in_equal  9.567175e-03      303      69
209        and_discontinued  9.945299e-03      276     109
210      company_recognizes  1.005958e-02      710     363
211                defaults  1.027979e-02      549     241
212                 raising  1.051244e-02      811     289
213         related_parties  1.164376e-02      289     181
214               covenants  1.220933e-02     4560    1334
215        negotiating_with  1.231684e-02      215      59
216             for_related  1.265269e-02      371      31
217             convertible  1.369356e-02     3317     908
218                 and_iii  1.374920e-02     2016     386
219                net_loss  1.526683e-02     4861    1439
220          administrative  1.532421e-02    10877    2555
221     company_experienced  1.585762e-02     1648     243
222           b_convertible  1.691292e-02      203      44
223                  year_#  2.089740e-02     9062    1729
224               a_waiver  2.302749e-02      462     163
225           default_under  2.329631e-02      498     240
226     financial_covenants  2.831332e-02     1902     642
227           negative_cash  3.395927e-02      563     223
228 reporting_comprehensive  3.414236e-02     1309       3
229                risk_not  3.445474e-02      845      54
230            its_computer  4.091666e-02     1348      25
231                 a_going  1.153621e-01      384     167
232              NT_before  7.594438e-01    12265    2845
Bias term:
       s0
-0.775231
*********************************************************
*                       2003                            *
*********************************************************
final_test       2003
alpha    1
dfmax    5000
year_window      5
refine   TRUE
input_dir        /usr2/brendano/10K/feat_bigrams/prune_30k/2003.relprune
input_ext        tf.num
more_nontext_data_file
nohist   FALSE
outfile          results.RData
final_test       2003
final_train      1998 1999 2000 2001 2002
dev_test         2002
```

dev_train 1998 1999 2000 2001

	name	beta	train_df	test_df
1	merger_is	-0.0867171086	264	44
2	exit_or	-0.0676941120	350	2015
3	and_plan	-0.0656337728	540	218
4	merger_agreement	-0.0630669832	649	166
5	computed_in	-0.0197309331	276	97
6	tender_offer	-0.0193800787	625	239
7	weather	-0.0175063047	1924	654
8	item_#a	-0.0173626911	9588	3364
9	mortgage_notes	-0.0154840023	264	92
10	of_merger	-0.0152163635	501	120
11	net_income	-0.0148558140	9005	2743
12	million_lower	-0.0139628041	352	216
13	not_represent	-0.0135362256	631	261
14	about_market	-0.0130696062	9671	3370
15	market_risk	-0.0117386449	10378	3456
16	properties	-0.0113520071	3423	1284
17	at_september	-0.0111701820	1013	300
18	the_objectives	-0.0110912132	276	92
19	distribution_system	-0.0110784007	292	104
20	rates	-0.0105704659	11334	3513
21	operating_partnership	-0.0098341607	308	84
22	adopted	-0.0092362281	5220	2644
23	rate	-0.0091182846	12110	3523
24	improved	-0.0090660688	5759	1895
25	real_estate	-0.0088953609	2930	1205
26	estate_taxes	-0.0080355105	442	179
27	rising	-0.0071997749	1367	537
28	risk	-0.0066660572	11607	3559
29	rental_rates	-0.0054922998	382	152
30	heating	-0.0053774340	397	131
31	reit	-0.0053243302	516	162
32	market	-0.0051901946	12434	3601
33	changes_in	-0.0046108831	10885	3459
34	accounting_for	-0.0043372062	7204	3104
35	distributions	-0.0043232033	2010	684
36	repurchases	-0.0041457567	1578	713
37	gains	-0.0040819104	6096	2396
38	tax_rate	-0.0035213153	6052	1904
39	year_s	-0.0034291557	1539	537
40	earnings	-0.0032030130	9545	2985
41	medium	-0.0019165239	788	266
42	rate_changes	-0.0018951947	2061	842
43	income_the	-0.0018740150	2487	879
44	declared_a	-0.0015001422	455	166
45	change_in	-0.0012850370	8283	3018
46	tenants	-0.0012653304	488	181
47	securities_and	-0.0011924497	6036	2000
48	units_and	-0.0010664539	939	393
49	reits	-0.0010409384	317	85
50	higher_operating	-0.0009809158	754	238
51	of_marketable	-0.0008085267	919	358
52	#table_of	-0.0004611335	515	1049
53	share	-0.0003841538	10347	3204
54	#_accounting	-0.0003830889	6382	2943
55	strong	-0.0002183169	3827	1289
56	improvements	-0.0002117556	5691	1906
57	noncompliance	0.0010012036	1277	288
58	distance	0.0010218446	506	140
59	the_assets	0.0010632734	4406	1968
60	wages	0.0011750112	1193	426
61	establishes_standards	0.0012065239	855	253
62	however_such	0.0016134770	355	105
63	raise	0.0017771416	2819	1118
64	subordinated_notes	0.0019204368	1531	491
65	debenture	0.0019382524	390	104

```
66          working_capital  0.0019919654     9697    2736
67              systems_and  0.0020206383     5955    1089
68               the_senior  0.0021700965     1418     549
69                 #_issues  0.0022065538     3365      37
70                 software  0.0022100410     7626    1441
71       additional_capital  0.0023673883     2485     844
72                divisions  0.0025109426     1086     353
73                  lenders  0.0030229984     2063     798
74                    issue  0.0030494113     6890    2301
75               additional  0.0031826504    12232    3537
76            increased_its  0.0042457382     1448     322
77                  vendors  0.0042721547     4750     936
78                    annum  0.0042902040     2217     725
79        company_completed  0.0044696052     3267     647
80     additional_financing  0.0046732147     2453     815
81           its_operations  0.0049303812     4465     875
82             covenants_in  0.0050437449      621     311
83                financing  0.0067447167    10921    3225
84                  will_be  0.0073289708    12098    3457
85        obtain_additional  0.0076883338     1514     607
86                 covenant  0.0081870535     1254     592
87             financing_or  0.0084219128      751     270
88            going_concern  0.0085160357      484     190
89                 computer  0.0085275530     7485    1230
90       acquired_companies  0.0086622013      903     298
91           administrative  0.0089292019    11441    3220
92                compliant  0.0090029178     3895      67
93                 an_event  0.0094823510      673     483
94      company_experienced  0.0095241175     1588     308
95              loss_before  0.0099190481     2743     924
96               in_default  0.0102094148      620     254
97                that_such  0.0118112577     4196    1217
98               its_future  0.0119816139     1359     313
99      financial_covenants  0.0122081639     2255     865
100               covenants  0.0132406441     5134    1843
101                  year_#  0.0155208637     9899    2315
102           no_assurance  0.0156143569     7593    2035
103               net_loss  0.0160402157     5494    1845
104             convertible  0.0192336928     3663    1145
105                risk_not  0.0224005159      816      55
106 reporting_comprehensive  0.0253704489     1206       4
107                a_waiver  0.0266453731      573     228
108          negative_cash  0.0268639576      704     235
109            for_related  0.0289037447      366      53
110           its_computer  0.0395397204     1326      20
111                a_going  0.1194823245      497     192
112              NT_before  0.7774560246    12850    3611
Bias term:
        s0
-0.6717631
******************************************************
*                      2004                          *
******************************************************
final_test       2004
alpha    1
dfmax    5000
year_window      5
refine   TRUE
input_dir        /usr2/brendano/10K/feat_bigrams/prune_30k/2004.relprune
input_ext        tf.num
more_nontext_data_file
nohist   FALSE
outfile          results.RData
final_test       2004
final_train      1999 2000 2001 2002 2003
dev_test         2003
dev_train        1999 2000 2001 2002
```

	name	beta	train_df	test_df
1	merger_agreement	-9.253670e-02	707	172
2	and_plan	-3.560503e-02	686	259
3	distribution_system	-3.358502e-02	344	126
4	exit_or	-2.843788e-02	2363	670
5	of_recognizing	-2.504935e-02	314	97
6	tender_offer	-2.365389e-02	760	266
7	computed_in	-2.298689e-02	318	102
8	or_modified	-2.233470e-02	1699	1404
9	no_impairment	-1.990374e-02	971	667
10	critical_accounting	-1.968700e-02	5514	3338
11	plan_assets	-1.804694e-02	647	597
12	net_income	-1.759636e-02	9989	2712
13	to_charge	-1.579750e-02	378	162
14	#_guarantor	-1.524322e-02	1663	681
15	york_stock	-1.458018e-02	572	286
16	properties	-1.413891e-02	4089	1398
17	units_and	-1.378495e-02	1174	436
18	the_proposed	-1.374947e-02	1242	475
19	expenditures_are	-1.313778e-02	1193	387
20	adopted	-1.292069e-02	7128	2346
21	had_entered	-1.242033e-02	471	153
22	share_repurchases	-1.218854e-02	449	195
23	tenants	-1.215365e-02	597	207
24	and_disclosure	-1.140107e-02	3067	1411
25	reporting_unit	-1.119757e-02	1086	766
26	amortization_as	-1.109202e-02	469	161
27	to_consider	-1.107527e-02	785	309
28	of_merger	-1.092437e-02	527	100
29	weather	-1.081682e-02	2224	744
30	market_information	-1.036596e-02	329	169
31	rates	-1.029926e-02	13069	3470
32	is_inherently	-9.716358e-03	366	197
33	and_regulatory	-9.676671e-03	2061	788
34	noncurrent	-9.636381e-03	733	337
35	audits	-9.416362e-03	751	445
36	gains	-9.147596e-03	7505	2279
37	with_gaap	-8.481255e-03	563	340
38	pension	-8.286042e-03	1518	908
39	of_premium	-8.243892e-03	425	162
40	interest_entities	-8.177975e-03	1399	1867
41	is_reduced	-7.778293e-03	754	323
42	shelf	-7.469224e-03	1139	507
43	tax_rate	-7.272860e-03	6826	1968
44	for_stockbased	-7.245190e-03	2219	872
45	that_expires	-7.234628e-03	442	215
46	s_accounting	-7.001447e-03	2371	972
47	led_by	-6.996992e-03	604	225
48	war	-6.990086e-03	873	604
49	property_management	-6.874473e-03	353	105
50	approval_of	-6.845676e-03	2469	870
51	contents	-6.533949e-03	1648	1378
52	of_fas	-6.533618e-03	613	186
53	which_represents	-6.206127e-03	1446	557
54	property_acquisitions	-6.159630e-03	433	122
55	improved	-6.142448e-03	6496	2191
56	as_extraordinary	-6.094940e-03	555	159
57	are_generally	-5.635949e-03	4506	1719
58	repayments	-5.520031e-03	2844	922
59	#_$	-5.395181e-03	8522	3256
60	and_higher	-5.377434e-03	3936	1242
61	fees_or	-5.321437e-03	360	151
62	by_operating	-5.311958e-03	5399	1672
63	disclosures	-5.131123e-03	11462	3239
64	disclosure_provisions	-4.951939e-03	1037	353
65	earnings_were	-4.873090e-03	622	213
66	behind	-4.847650e-03	538	210

67	current_liability	-4.816274e-03	321	152
68	contributed	-4.706156e-03	5542	1795
69	corrections_sfas	-4.550561e-03	548	123
70	the_terrorist	-4.461658e-03	541	178
71	flows_and	-4.451467e-03	2786	1280
72	limited_liability	-4.449849e-03	564	212
73	medium	-4.355055e-03	955	303
74	recorded_net	-4.218513e-03	994	377
75	earnings_are	-4.143227e-03	868	296
76	utility	-4.102553e-03	1581	484
77	#%_higher	-3.908107e-03	917	295
78	price_increase	-3.852616e-03	371	157
79	counterparties	-3.816431e-03	852	351
80	announced_that	-3.728484e-03	1853	632
81	fair_values	-3.663089e-03	2948	1242
82	excluding	-3.573584e-03	7379	2159
83	real_estate	-3.569725e-03	3657	1272
84	paper	-3.562841e-03	2633	792
85	repurchases	-3.514711e-03	2086	781
86	fin	-3.429723e-03	1834	1763
87	estate_taxes	-3.423463e-03	545	200
88	and_represent	-3.314007e-03	341	134
89	value_as	-3.158517e-03	1401	633
90	military	-3.120594e-03	887	474
91	%	-3.092627e-03	4997	2335
92	regulated	-3.092164e-03	1170	483
93	will_receive	-3.084759e-03	1304	456
94	incentive	-3.066389e-03	2917	1136
95	changes_in	-2.828784e-03	12617	3438
96	annual_impairment	-2.810021e-03	905	548
97	proposed	-2.753952e-03	2791	1098
98	revenues_$	-2.710746e-03	1411	678
99	be_classified	-2.635660e-03	939	535
100	lower_interest	-2.634988e-03	3113	1345
101	estimates	-2.594425e-03	9707	3410
102	assumptions	-2.470868e-03	7963	3231
103	borrowing_capacity	-2.464528e-03	1474	530
104	prepared_in	-2.362675e-03	2695	1462
105	of_gas	-2.320051e-03	382	136
106	conditions_including	-2.287644e-03	1069	433
107	eitf	-2.287234e-03	2391	1258
108	types_of	-2.283895e-03	3690	1556
109	rate	-2.235566e-03	13461	3484
110	securities_and	-2.221942e-03	7147	2139
111	political	-2.210311e-03	3552	1273
112	generally	-2.107343e-03	11665	3436
113	program_under	-2.077609e-03	337	134
114	of_marketable	-1.997135e-03	1160	362
115	from_operating	-1.990421e-03	3988	1299
116	employee_stock	-1.944049e-03	2467	986
117	tax_returns	-1.868181e-03	686	415
118	taxes_in	-1.803627e-03	2261	930
119	the_objectives	-1.753175e-03	324	96
120	considers	-1.737790e-03	2708	1020
121	in_other	-1.728727e-03	7386	2407
122	liability	-1.714429e-03	8642	3032
123	improvement	-1.709493e-03	5237	1917
124	distributions	-1.609478e-03	2307	766
125	both	-1.594768e-03	11732	3351
126	senior_unsecured	-1.540231e-03	568	288
127	conditions	-1.535713e-03	11672	3394
128	the_responsibility	-1.512904e-03	469	136
129	the_transactions	-1.491270e-03	856	343
130	plan_of	-1.468661e-03	1231	411
131	circumstances	-1.367582e-03	7266	3027
132	earnings_increased	-1.365379e-03	319	99
133	conditions_and	-1.360750e-03	5007	1821

134	termination_fee	-1.331582e-03	322	128
135	audit	-1.319025e-03	2043	1188
136	interstate	-1.317944e-03	405	170
137	recover	-1.307602e-03	2287	1028
138	insurance	-1.247860e-03	7376	2691
139	included_$#	-1.227299e-03	3150	1049
140	credit_ratings	-1.158694e-03	746	377
141	annual	-1.047376e-03	10788	3245
142	loss_related	-1.029871e-03	401	166
143	in_net	-1.003138e-03	7542	2210
144	change_in	-9.792423e-04	10078	3062
145	earnings	-9.495147e-04	10814	3055
146	were_partially	-9.007162e-04	3803	1344
147	and_changes	-8.831634e-04	4194	1551
148	reconciliation	-8.449515e-04	886	513
149	were_higher	-7.991430e-04	1003	384
150	increasing_our	-7.358100e-04	760	412
151	amortizing_goodwill	-7.319483e-04	381	194
152	statements_contained	-7.314108e-04	2614	702
153	longer_amortized	-6.387296e-04	589	287
154	rates_have	-6.291265e-04	816	322
155	determining	-5.666804e-04	3816	1851
156	million_investment	-5.435331e-04	418	131
157	estimation	-4.650395e-04	1074	675
158	partnership	-4.544044e-04	2205	728
159	contractual	-4.464620e-04	6782	3315
160	unit	-4.235192e-04	5614	2010
161	more_than	-3.963649e-04	6382	2570
162	differences	-3.849402e-04	4836	2029
163	#year	-3.744004e-04	1528	630
164	most	-2.791601e-04	10050	3094
165	increases	-2.771980e-04	11751	3185
166	when	-2.510992e-04	11349	3421
167	index	-2.344499e-04	1265	512
168	aware_that	-2.221671e-04	456	152
169	gas	-2.053562e-04	1631	584
170	returns	-1.920823e-04	4076	1979
171	settlement	-1.790810e-04	5152	1925
172	than_in	-1.754094e-04	2597	980
173	eitf_no	-1.565485e-04	479	273
174	of_an	-1.542122e-04	9508	2924
175	change	-1.326184e-04	12196	3403
176	rental_rates	-1.226918e-04	467	164
177	funds_for	-9.033692e-05	1582	485
178	political_conditions	-4.896766e-05	549	250
179	cumulative_effect	-6.499596e-06	2655	1037
180	than	-4.832743e-06	13306	3513
181	is_due	5.688940e-05	5758	1585
182	company_is	1.149113e-04	8497	1839
183	loss	2.059716e-04	12876	3439
184	not_experienced	2.079519e-04	1437	309
185	comparable_period	2.354342e-04	888	234
186	covenant	3.017770e-04	1680	601
187	prime_rate	3.716545e-04	3481	917
188	$#_the	3.885348e-04	5181	1296
189	date	4.208726e-04	12343	3332
190	an_event	5.582011e-04	1085	536
191	increased_its	7.930066e-04	1418	370
192	operating_loss	8.663571e-04	5072	1541
193	broadband	9.389451e-04	591	196
194	value_assigned	9.509518e-04	344	80
195	borrowings_to	9.895949e-04	1003	246
196	continue_as	9.906254e-04	815	238
197	will_include	1.177632e-03	675	203
198	telecommunications	1.180062e-03	2723	711
199	a_right	1.357026e-03	419	176
200	s_working	1.484911e-03	1343	231

201	capital_resources	1.673219e-03	12997	3266	
202	consulting	1.954176e-03	4413	1268	
203	warrants_to	1.960400e-03	1578	382	
204	fiber	2.014494e-03	660	201	
205	its_operations	2.024326e-03	4260	814	
206	sales_for	2.078157e-03	4501	1238	
207	aggregate_of	2.123205e-03	2326	751	
208	operations_year	2.193851e-03	1445	332	
209	stage	2.279644e-03	2192	633	
210	#_issue	2.371777e-03	2225	35	
211	page	2.416083e-03	10085	1321	
212	certain_covenants	2.459788e-03	861	230	
213	event_the	2.572350e-03	1543	474	
214	vendors	2.698172e-03	4810	1046	
215	of_components	2.771548e-03	637	209	
216	lower_margins	2.809077e-03	1162	316	
217	lenders	2.899604e-03	2550	800	
218	$#_of	2.901684e-03	5481	1441	
219	startup	2.904154e-03	2582	520	
220	not_incurred	2.907578e-03	457	122	
221	addresses_financial	3.049411e-03	1825	263	
222	seeking	3.071019e-03	2631	910	
223	solutions	3.214838e-03	3286	1041	
224	#_page	3.313545e-03	8974	850	
225	consideration_paid	4.112862e-03	512	118	
226	divisions	4.151400e-03	1237	381	
227	restructure_the	4.225536e-03	326	87	
228	the_preferred	4.313569e-03	1273	371	
229	of_client	4.350935e-03	336	85	
230	systems_and	4.352836e-03	5846	1173	
231	a_senior	4.440744e-03	660	245	
232	readiness_disclosure	4.619463e-03	417	0	
233	business_plan	4.732739e-03	1117	385	
234	delisting	5.056435e-03	371	98	
235	its_computer	5.072887e-03	699	15	
236	#_readiness	5.115509e-03	1483	2	
237	that_such	5.160666e-03	4521	1257	
238	company_completed	5.194871e-03	3102	558	
239	smallcap_market	5.239657e-03	317	98	
240	of_technology	5.394975e-03	1223	430	
241	or_equity	5.537031e-03	1821	578	
242	#_problems	5.691448e-03	1029	3	
243	working_capital	5.697140e-03	10577	2732	
244	year_#	5.914709e-03	9945	2587	
245	its_control	6.052754e-03	324	87	
246	apart	6.055380e-03	624	151	
247	purchase_method	6.342411e-03	3145	407	
248	combinations_initiated	6.478802e-03	1571	90	
249	in_default	6.526450e-03	803	232	
250	operating_losses	6.559834e-03	4356	1241	
251	placements	6.628469e-03	1049	272	
252	loss_before	6.827991e-03	3226	862	
253	raise	6.857891e-03	3570	1168	
254	are_secured	6.899598e-03	1518	487	
255	additional_capital	6.929091e-03	2922	870	
256	waiver	7.257861e-03	1046	318	
257	problems	7.261804e-03	5228	1048	
258	continued_listing	7.378152e-03	342	78	
259	annum	7.386584e-03	2538	733	
260	the_senior	7.448566e-03	1729	577	
261	measure_those	7.455747e-03	650	17	
262	waiver_of	7.457118e-03	376	95	
263	the_automotive	7.571329e-03	452	157	
264	debenture	7.710885e-03	419	142	
265	internal_systems	8.141270e-03	1193	61	
266	computer	8.308412e-03	6880	1228	
267	y#k	8.600813e-03	980	7	

```
268           transitioned  8.606354e-03      351     149
269              a_waiver  8.700817e-03      718     206
270           however_such  8.898405e-03      396     120
271   additional_financing  9.305315e-03     2905     791
272               will_be  9.461725e-03    13231    3432
273           financing_or  1.020321e-02      916     282
274              #_problem  1.055845e-02      946       6
275           were_charged  1.077113e-02      369     116
276         administrative  1.087141e-02    12488    3177
277          material_year  1.089607e-02      368       1
278          accounting_be  1.128468e-02      493      20
279                loss_#%  1.151209e-02     1276      98
280         fund_operating  1.159907e-02      332      90
281           no_assurance  1.190720e-02     8114    2018
282              financing  1.232467e-02    12187    3248
283      capital_intensive  1.297993e-02      344     131
284               net_loss  1.408367e-02     6433    1692
285        experienced_any  1.457808e-02     1078     183
286             its_future  1.470206e-02     1336     306
287      and_extraordinary  1.480981e-02     1778     144
288     acquired_companies  1.500344e-02     1074     309
289          generate_cash  1.537209e-02      644     320
290    financial_covenants  1.543274e-02     2761     884
291              merger_in  1.664720e-02      398      89
292       corporate_office  1.691843e-02      580     213
293             trade_show  1.805008e-02      424     110
294            be_disposed  1.816137e-02     2970     361
295 establishes_accounting  1.816565e-02     1700      79
296            monitor_its  1.898805e-02      394      49
297               distance  1.909710e-02      576     149
298          noncompliance  1.921460e-02     1382     324
299          company_wrote  2.034660e-02      427      88
300                  #_y#k  2.041465e-02      408       1
301              compliant  2.085144e-02     2816     106
302              bid_price  2.255447e-02      414      84
303               #_issues  2.260935e-02     2540      40
304     additional_working  2.619121e-02      309      75
305          negative_cash  2.849296e-02      841     215
306         experienced_no  3.139331e-02      468      31
307       significant_year  3.232652e-02      426       6
308          going_concern  4.904431e-02      626     128
309                a_going  6.900513e-02      642     134
310              NT_before  7.762639e-01    14000    3558
Bias term:
        s0
-0.7627311
******************************************************
*                       2005                         *
******************************************************
final_test       2005
alpha    1
dfmax    5000
year_window      5
refine   TRUE
input_dir        /usr2/brendano/10K/feat_bigrams/prune_30k/2005.relprune
input_ext        tf.num
more_nontext_data_file
nohist   FALSE
outfile          results.RData
final_test       2005
final_train      2000 2001 2002 2003 2004
dev_test         2004
dev_train        2000 2001 2002 2003
                     name          beta train_df test_df
1         merger_agreement -8.530324e-02      732     178
2             the_proposed -1.829192e-02     1505     523
3           for_stockbased -1.787200e-02     3079    1520
```

4	distribution_system	-1.649986e-02	413	123
5	critical_accounting	-1.333385e-02	8850	3263
6	with_exit	-1.304335e-02	2986	190
7	for_guarantees	-1.228275e-02	2437	135
8	exit_or	-1.157525e-02	3030	203
9	final_settlement	-1.071452e-02	388	157
10	plan_assets	-9.996786e-03	1204	586
11	net_income	-9.498014e-03	10916	2815
12	rates	-9.263198e-03	14328	3397
13	guarantees_issued	-8.844123e-03	1849	55
14	ratings	-8.467489e-03	1999	726
15	assumptions	-7.282053e-03	10239	3177
16	multiple_deliverables	-6.957407e-03	1075	277
17	gains	-6.289050e-03	8642	2224
18	changes_in	-6.243706e-03	13938	3360
19	share_repurchase	-6.219554e-03	956	326
20	of_pension	-5.589342e-03	503	257
21	proposed	-5.270759e-03	3438	1182
22	estimates	-4.828645e-03	11521	3334
23	tax_rate	-4.597773e-03	7622	2065
24	no_impairment	-4.449934e-03	1622	672
25	change	-4.286703e-03	13593	3321
26	exceeded	-4.182103e-03	3518	1154
27	reporting_unit	-4.046138e-03	1850	821
28	billion	-3.935099e-03	3310	1066
29	the_fair	-3.777314e-03	10329	2929
30	#_$	-3.630397e-03	10664	3228
31	plan_of	-3.500023e-03	1444	392
32	eitf	-3.473372e-03	3595	1131
33	#_guarantor	-3.039540e-03	2344	118
34	$	-2.908691e-03	11629	3337
35	liability	-2.795786e-03	10522	2935
36	in_other	-2.763359e-03	8666	2301
37	policies	-2.524490e-03	11392	3355
38	repurchases	-2.436358e-03	2637	869
39	and_plan	-1.786384e-03	841	282
40	improved	-1.654426e-03	7581	2196
41	by_higher	-1.587177e-03	2730	903
42	are_generally	-1.459446e-03	5592	1750
43	more_than	-1.431101e-03	7961	2572
44	earnings	-1.405844e-03	12052	3094
45	regulatory	-1.386367e-03	7225	2111
46	than_#	-1.347551e-03	6232	2319
47	calculated	-1.231637e-03	4710	1762
48	contents	-1.231125e-03	3010	1546
49	our_estimate	-1.192765e-03	1339	713
50	and_disclosure	-1.133867e-03	4406	1108
51	excluding	-1.089150e-03	8347	2176
52	most	-1.081846e-03	11312	3076
53	lower_interest	-9.747953e-04	4154	992
54	insurance	-7.008752e-04	8979	2692
55	determining	-6.643226e-04	5232	2037
56	s_pension	-4.787182e-04	470	213
57	settlement	-4.674784e-04	6379	2009
58	%	-2.829086e-04	6766	2494
59	paper	-2.571853e-04	3038	804
60	#_amendment	-2.160360e-04	2590	183
61	of_merger	-1.750055e-04	534	122
62	returns	-1.271438e-04	5640	1944
63	war	-1.124919e-04	1462	562
64	approved_the	-8.805797e-05	1153	383
65	estimated	-6.756120e-05	12077	3258
66	reconciliation	-4.335743e-05	1316	542
67	contractual	-4.216737e-05	9427	3265
68	circumstances	-2.628550e-05	9502	2988
69	approval_of	-1.917056e-05	2988	954
70	readiness	1.335710e-04	638	39

```
71              the_senior  5.133534e-04     2048     623
72         purchase_method  5.492002e-04     3295     322
73                distance  6.314170e-04      618     156
74   business_combinations  8.347244e-04     3991     710
75     financial_covenants  8.992464e-04     3258     929
76         management_will  9.665982e-04      623     193
77                 waivers  1.560882e-03      616     185
78                    page  1.732081e-03     9084    1039
79                  waived  1.872655e-03      853     205
80               financing  2.410173e-03    13328    3180
81      and_administrative  2.584254e-03    12638    2848
82           business_plan  3.111643e-03     1367     375
83        operating_losses  3.239022e-03     5016    1197
84                 a_going  3.760745e-03      685     145
85                  waiver  5.046813e-03     1211     282
86            combinations  5.107528e-03     4246     800
87          administrative  5.186474e-03    13411    3072
88    increased_borrowings  5.614949e-03      708     101
89                  #_page  6.061616e-03     7676     554
90                 #_issue  6.561822e-03      789      28
91           noncompliance  6.891102e-03     1128     336
92              in_default  6.949802e-03      938     243
93                a_waiver  7.131471e-03      826     193
94              its_future  7.359563e-03     1317     292
95                problems  7.716248e-03     4677    1075
96           no_assurance  7.813711e-03     8364    1970
97            warrants_to  7.865197e-03     1710     366
98            monitor_its  8.732271e-03      404      43
99               net_loss  9.355186e-03     7143    1539
100 establishes_accounting  1.005100e-02     1424      58
101        working_capital  1.022185e-02    11411    2624
102                    y#k  1.104440e-02      495       3
103           be_disposed  1.410572e-02     3251     217
104             compliant  1.481818e-02      897     132
105               loss_#%  1.688386e-02     1098      63
106 combinations_initiated  1.696145e-02     1661      19
107          negative_cash  1.740384e-02      962     190
108   additional_financing  1.778657e-02     3273     759
109       internal_systems  1.826273e-02      612      54
110         experienced_no  2.018281e-02      486      32
111              bid_price  2.369008e-02      452      83
112           #_readiness  2.535688e-02      425       3
113          #_compliance  3.263143e-02      875     171
114              #_issues  4.443332e-02      973      47
115          going_concern  1.006828e-01      668     136
116              NT_before  8.164359e-01    15034    3474
Bias term:
        s0
-0.6422074
******************************************************
*                       2006                         *
******************************************************
final_test       2006
alpha    1
dfmax    5000
year_window      5
refine   TRUE
input_dir        /usr2/brendano/10K/feat_bigrams/prune_30k/2006.relprune
input_ext        tf.num
log      TRUE
more_nontext_data_file
nohist   FALSE
outfile          results.RData
final_test       2006
final_train      2001 2002 2003 2004 2005
dev_test         2005
dev_train        2001 2002 2003 2004
```

	name	beta	train_df	test_df
1	merger_agreement	-1.220597e-01	770	148
2	the_proposed	-2.336379e-02	1813	389
3	exit_or	-1.961737e-02	3229	156
4	for_guarantees	-1.929499e-02	2571	91
5	guarantees_issued	-1.537776e-02	1904	33
6	poor_s	-1.413177e-02	1199	355
7	is_terminated	-1.399563e-02	558	138
8	consummation	-1.374788e-02	1099	230
9	of_merger	-1.365815e-02	554	130
10	distribution_system	-1.280600e-02	472	89
11	net_income	-1.206735e-02	12036	2702
12	multiple_deliverables	-1.180579e-02	1352	264
13	#_amendment	-1.071218e-02	2743	110
14	goodwill_amortization	-9.982982e-03	3000	59
15	upon_termination	-9.254841e-03	475	119
16	rates	-8.803012e-03	15541	3231
17	for_stockbased	-8.276976e-03	4591	1346
18	not_completed	-8.275528e-03	654	106
19	no_impairment	-8.071264e-03	2284	602
20	gains	-7.494588e-03	9739	2131
21	in_different	-7.090201e-03	1104	263
22	billion	-6.823156e-03	3952	1073
23	merger	-6.451867e-03	4467	824
24	ratings	-6.382549e-03	2547	725
25	termination_fee	-5.648338e-03	492	144
26	final_settlement	-5.308107e-03	507	146
27	customary	-4.917466e-03	2330	714
28	share_repurchase	-4.664707e-03	1142	375
29	by_an	-4.329039e-03	8586	2094
30	rate	-4.242222e-03	15652	3234
31	war	-4.004797e-03	1997	331
32	tender	-3.661566e-03	1600	310
33	changes_in	-3.640353e-03	15214	3205
34	earnings	-3.447789e-03	13348	2898
35	tax_rate	-3.298224e-03	8520	1997
36	the_fair	-3.172943e-03	12237	2854
37	amortization_as	-3.134007e-03	612	104
38	%	-2.866777e-03	8631	2445
39	subsidiary_of	-2.830435e-03	3890	887
40	disclosure_provisions	-2.740943e-03	1509	60
41	practices	-2.691411e-03	5255	1169
42	announcement_of	-2.635308e-03	673	116
43	with_gaap	-2.471478e-03	1181	423
44	across	-2.248759e-03	4126	1148
45	maintenance_revenue	-2.214731e-03	586	110
46	had_entered	-1.868056e-03	608	126
47	improved	-1.786447e-03	8690	2035
48	by_operating	-1.783688e-03	7002	1638
49	as_lower	-1.654965e-03	973	187
50	insurance	-1.559866e-03	10650	2383
51	changes	-1.542665e-03	15516	3240
52	such_that	-1.541824e-03	1667	461
53	more_than	-1.516374e-03	9616	2436
54	annual	-1.496631e-03	13929	3020
55	repurchases	-1.495655e-03	3181	925
56	on_plan	-1.343555e-03	1478	479
57	rate_increases	-1.180104e-03	1533	438
58	lower_interest	-1.151140e-03	4856	516
59	million_shares	-1.133619e-03	4235	949
60	plan_assets	-1.015421e-03	1774	543
61	s_pension	-1.006652e-03	672	185
62	#_$	-7.525399e-04	12757	3072
63	with_exit	-7.462453e-04	3167	154
64	consummation_of	-1.947461e-04	1034	218
65	policies	-1.518534e-05	13847	3191
66	delisted	2.920512e-05	561	44

```
67                  #%_per  1.748336e-04     3309     744
68          years_beginning  1.957391e-04     4578    1361
69                  $#_from  3.174019e-04     4357     863
70                   is_due  3.424793e-04     7026    1461
71          ending_december  3.714620e-04     2427     599
72               in_default  6.046037e-04     1064     228
73                financing  6.304162e-04    14447    3011
74  establishes_accounting  6.703853e-04     1058      45
75                   lender  7.974608e-04     2601     593
76     financial_covenants  8.402379e-04     3777     876
77                debenture  9.726663e-04      547     116
78                   $#_for  1.025146e-03     8249    1643
79                    fiber  1.546667e-03      870     162
80            goodwill_will  1.786705e-03      781      50
81                   #_page  2.107408e-03     6227     337
82                 a_waiver  2.311206e-03      900     172
83          inventory_costs  2.574790e-03     1218     712
84                 covenant  2.775305e-03     2444     560
85          supersedes_sfas  3.257020e-03      821      18
86            business_plan  3.519738e-03     1605     308
87           operating_loss  3.535225e-03     6611    1434
88          management_will  4.291314e-03      724     156
89               the_senior  4.578007e-03     2403     584
90           administrative  4.608883e-03    14334    2875
91            for_longlived  5.218785e-03     2767     271
92            accounting_be  5.306448e-03      523       7
93                assembled  5.434160e-03      818     115
94               its_future  6.440989e-03     1359     213
95             no_assurance  7.028874e-03     8949    1618
96               trade_show  7.091767e-03      534     108
97              warrants_to  9.789740e-03     1819     337
98                  working  1.048482e-02    12808    2577
99    additional_financing  1.216526e-02     3603     614
100                   raise  1.232084e-02     5024    1035
101               net_loss  1.340893e-02     7720    1402
102             be_disposed  1.487853e-02     3380     175
103                  waived  1.497252e-02      926     181
104  business_combinations  1.578750e-02     4548     541
105       profit_decreased  2.154973e-02     1122     167
106          negative_cash  2.324006e-02     1040     151
107              bid_price  3.681609e-02      499      48
108 combinations_initiated  4.065514e-02     1679       8
109                 a_going  4.905490e-02      753     114
110          going_concern  6.415892e-02      734     107
111              NT_before  7.816477e-01    16084    3306
Bias term:
        s0
-0.8030612
```

unigrams (run26_nt_no_penalty/model_reports.txt)

```
******************************************************
*                       2001                         *
******************************************************
hi
final_test       2001
alpha    1
dfmax    5000
year_window      5
refine   TRUE
input_dir        /usr2/brendano/10K/feat3countsel/fc5_relprune/2001.relprune
input_ext        tf.num
```

nohist	FALSE			
outfile	results.RData			
final_test	2001			
final_train	1996 1997 1998 1999 2000			
dev_test	2000			
dev_train	1996 1997 1998 1999			

	name	beta	train_df	test_df
1	mardi	-5.765737e-01	6	0
2	swine	-4.122798e-01	8	1
3	#%a	-4.099695e-01	6	3
4	selfconstructed	-3.791738e-01	7	0
5	recertification	-2.709669e-01	8	2
6	subsidizes	-2.694879e-01	5	0
7	otto	-2.580502e-01	5	0
8	thepage	-2.502449e-01	6	0
9	longtail	-2.345215e-01	5	3
10	fortis	-2.077593e-01	9	4
11	transmittal	-1.824058e-01	7	4
12	fahrenheit	-1.767818e-01	13	2
13	setaside	-1.389664e-01	5	1
14	thermoforming	-1.309977e-01	5	0
15	validly	-1.221395e-01	27	6
16	atg	-1.116375e-01	5	4
17	collateralbased	-1.015218e-01	5	2
18	unanimously	-9.336370e-02	76	12
19	cigar	-9.309009e-02	11	0
20	osprey	-8.968859e-02	6	3
21	irradiation	-8.371068e-02	8	0
22	xl	-6.401121e-02	25	5
23	samelocation	-6.356514e-02	6	1
24	hsr	-6.215076e-02	14	1
25	interstatejohnson	-5.545086e-02	6	0
26	creditwatch	-4.871986e-02	10	2
27	#d#	-4.578102e-02	31	6
28	midatlantic	-4.208049e-02	105	27
29	tic	-4.026387e-02	9	2
30	discs	-3.744190e-02	33	9
31	murex	-3.691124e-02	5	1
32	sendout	-3.313662e-02	11	0
33	overpayment	-3.232290e-02	30	7
34	hose	-3.153410e-02	16	5
35	cobalt	-3.062514e-02	21	4
36	hcv	-2.742572e-02	5	4
37	lime	-2.677628e-02	17	4
38	reit	-2.578046e-02	401	107
39	sub	-2.552820e-02	126	26
40	surviving	-2.250692e-02	189	55
41	precast	-1.994597e-02	7	1
42	stockbased	-1.917541e-02	835	174
43	mgt	-1.601509e-02	7	0
44	scmc	-1.588204e-02	5	0
45	merger	-1.570662e-02	2714	750
46	fiserv	-1.560178e-02	17	4
47	nearly	-1.441119e-02	917	217
48	lifo	-1.350879e-02	256	49
49	weather	-1.312216e-02	1644	372
50	properties	-1.279207e-02	2816	668
51	nareit	-1.221309e-02	222	56
52	unitholders	-1.100340e-02	150	36
53	competi	-1.098745e-02	6	0
54	income	-1.065095e-02	10921	2571
55	measure	-1.063822e-02	2109	683
56	unsecured	-1.001678e-02	2518	531
57	vote	-9.708213e-03	952	254
58	tender	-9.660574e-03	734	220
59	revenuebased	-8.552035e-03	7	4
60	positive	-7.825054e-03	2612	660

61	ffo	-7.809415e-03	234	65	
62	massmarket	-7.323450e-03	16	4	
63	ldc	-6.948385e-03	28	4	
64	distributions	-6.613998e-03	1621	375	
65	cobank	-6.557202e-03	6	1	
66	estate	-5.713806e-03	2405	616	
67	rate	-5.465325e-03	10025	2495	
68	flavoring	-5.449496e-03	6	0	
69	improved	-5.097293e-03	5157	1093	
70	industrial	-5.015256e-03	2159	485	
71	real	-4.883637e-03	2810	730	
72	awarded	-4.700088e-03	694	164	
73	warmer	-4.131044e-03	186	37	
74	approval	-4.072702e-03	2986	832	
75	virtue	-4.052456e-03	140	38	
76	rising	-3.523810e-03	1054	328	
77	declared	-3.516497e-03	1580	346	
78	increases	-3.364070e-03	9072	2194	
79	longterm	-3.261766e-03	6899	1750	
80	isf	-3.225606e-03	7	0	
81	fixedrate	-2.536630e-03	594	259	
82	adopted	-2.264644e-03	3535	1155	
83	partnership	-2.033492e-03	1560	347	
84	repayments	-1.683966e-03	1728	506	
85	bulk	-1.534801e-03	590	137	
86	rates	-1.403893e-03	8836	2427	
87	intrinsic	-1.255589e-03	259	55	
88	average	-8.443223e-04	8147	2076	
89	$	-8.403026e-04	5357	1524	
90	taxexempt	-4.718093e-04	692	168	
91	dividend	-3.367520e-04	3121	727	
92	business	3.998012e-06	10297	2463	
93	accounts	1.842561e-04	7626	1884	
94	compliant	3.593930e-04	3832	68	
95	can	4.041991e-04	8833	2099	
96	discontinuation	6.733801e-04	356	99	
97	iii	8.905661e-04	3509	691	
98	comprehensive	9.030535e-04	3298	833	
99	expenses	1.066159e-03	10917	2562	
100	development	1.387706e-03	8042	1969	
101	debenture	1.573053e-03	348	77	
102	receivable	1.787096e-03	6462	1642	
103	efforts	1.899094e-03	6745	1514	
104	writedown	1.943300e-03	1629	479	
105	stage	1.970332e-03	1278	379	
106	obtain	2.258076e-03	4320	1185	
107	material	2.455961e-03	9476	2211	
108	forwardlooking	2.561788e-03	6079	1760	
109	covenant	2.950940e-03	823	246	
110	goodwill	2.953084e-03	3202	1203	
111	forgo	3.323200e-03	12	7	
112	purchasing	3.326635e-03	2510	654	
113	could	3.430676e-03	8944	2186	
114	consulting	3.585209e-03	2702	783	
115	assurance	3.649264e-03	6649	1425	
116	personnel	3.879632e-03	5746	1495	
117	enterprise	4.043060e-03	2357	376	
118	loss	4.229830e-03	9346	2341	
119	not	4.474488e-03	10950	2586	
120	duplicative	4.523418e-03	172	42	
121	date	4.666779e-03	8490	2090	
122	infrastructure	4.786184e-03	2477	801	
123	private	5.061548e-03	5564	1516	
124	amendment	5.063292e-03	1270	515	
125	financing	5.149368e-03	8907	2256	
126	report	5.303918e-03	6300	1704	
127	deficit	5.724263e-03	1246	416	

```
128           change#  5.744299e-03        8       3
129          warrants  6.452562e-03     2209     664
130          software  8.666810e-03     6269    1048
131   securitizations  9.146607e-03      176     124
132             china  9.281686e-03      586     180
133               mmi  1.061676e-02       12       7
134              cone  1.102859e-02        7       3
135          computer  1.107771e-02     6525     927
136        additional  1.166460e-02    10346    2461
137             raise  1.480971e-02     1732     676
138           raising  1.488614e-02      653     230
139       convertible  1.498356e-02     2869     777
140       forbearance  1.724173e-02       81      29
141       negotiating  1.814038e-02      831     209
142        recruiting  1.929728e-02      763     321
143             klein  2.132132e-02        8       1
144         covenants  2.247751e-02     3965    1004
145    administrative  2.322512e-02     9765    2310
146               usd  2.763032e-02       45      19
147              font  2.905312e-02        6       2
148          defaults  3.299152e-02      454     133
149        purposeful  3.808302e-02        5       2
150         westbrook  4.047250e-02        7       1
151            livery  4.260337e-02        5       2
152             going  4.404587e-02      696     256
153           starter  4.681170e-02       13       3
154             aruba  5.390673e-02        5       2
155    expenserelated  5.839337e-02        7       1
156        gainonsale  1.103300e-01        5       1
157         heapleach  1.229721e-01        5       0
158               ive  1.265774e-01        5       1
159          elsinore  2.540949e-01        5       2
160         NT_before  8.076907e-01    11075    2596
Bias term:
        s0
-0.6198935
******************************************************
*                        2002                        *
******************************************************
hi
final_test       2002
alpha    1
dfmax    5000
year_window      5
refine   TRUE
input_dir        /usr2/brendano/10K/feat3countsel/fc5_relprune/2002.relprune
input_ext        tf.num
nohist   FALSE
outfile          results.RData
final_test       2002
final_train      1997 1998 1999 2000 2001
dev_test         2001
dev_train        1997 1998 1999 2000
                 name          beta train_df test_df
1               mardi -0.5196468999        6       0
2              havent -0.4260958013        5       1
3               swine -0.3956668634        7       4
4     selfconstructed -0.2969237078        7       1
5              onpeak -0.2892173094        5       1
6                otto -0.1874591436        5       0
7           fullprice -0.1666080954       10       7
8     recertification -0.1353552662        9       3
9             validly -0.1313793949       32      13
10             fortis -0.1102457419       13       4
11        transmittal -0.1054962152       10       4
12     indefinitelife -0.0933651275        5       8
13        unanimously -0.0878665519       84      18
```

14	professions	-0.0858565757	13	6
15	cigar	-0.0805761227	10	2
16	ratepayers	-0.0681181397	29	7
17	interstatejohnson	-0.0639916068	5	1
18	accrediting	-0.0595159823	15	5
19	samelocation	-0.0585211453	6	0
20	fahrenheit	-0.0549502409	12	5
21	gilroy	-0.0493821497	5	3
22	hose	-0.0482885498	19	8
23	irradiation	-0.0451021386	8	1
24	psychology	-0.0430066497	11	6
25	scholarships	-0.0355336807	6	1
26	biostudies	-0.0329161072	5	4
27	cobank	-0.0321669581	5	3
28	hac	-0.0317125955	5	1
29	graduates	-0.0302635764	16	3
30	surviving	-0.0299857383	227	67
31	termdebt	-0.0292890678	13	4
32	collateralbased	-0.0271166537	7	1
33	juice	-0.0244592507	47	15
34	enact	-0.0233723301	60	36
35	nareit	-0.0222474404	271	50
36	doctoral	-0.0221765125	7	2
37	reit	-0.0213410753	487	105
38	reimaging	-0.0204699831	13	3
39	hcv	-0.0197364733	9	7
40	propulsion	-0.0190231500	28	12
41	murex	-0.0155684958	6	2
42	lime	-0.0153840474	19	5
43	rate	-0.0149960632	11310	2759
44	repurchases	-0.0137910127	1248	466
45	#d#	-0.0135995541	34	9
46	soybeans	-0.0132112520	15	4
47	weather	-0.0131873436	1814	438
48	ffo	-0.0123610286	290	61
49	rates	-0.0121651901	10202	2734
50	properties	-0.0115245997	3117	842
51	income	-0.0115163361	12109	2823
52	merger	-0.0112927017	3180	785
53	distributions	-0.0104276828	1857	458
54	warmer	-0.0076977576	192	52
55	tenants	-0.0075133110	443	118
56	temperatures	-0.0071848834	137	45
57	principles	-0.0070229026	2717	1962
58	tender	-0.0068011426	915	271
59	apartment	-0.0067817016	229	47
60	setaside	-0.0061602191	6	0
61	rising	-0.0060916017	1212	355
62	repayments	-0.0059025853	2047	617
63	increases	-0.0055844932	10112	2411
64	unsecured	-0.0053672423	2768	658
65	market	-0.0050745869	11512	2835
66	fiserv	-0.0049753964	20	3
67	dekalb	-0.0047963182	8	2
68	sendout	-0.0047065396	9	1
69	regulatory	-0.0046255171	4633	1301
70	hsr	-0.0044174535	15	1
71	strong	-0.0040228274	3484	916
72	derivative	-0.0039935987	5222	1721
73	sub	-0.0037759701	143	33
74	gains	-0.0031444774	5324	1548
75	excluding	-0.0030335579	5722	1590
76	instruments	-0.0027418901	6781	2219
77	derivatives	-0.0026376100	2821	730
78	remand	-0.0022300122	21	9
79	board	-0.0021506787	7554	2367
80	adopted	-0.0020218041	4214	1676

81	authorization	-0.0020084830	784	272	
82	reinvestment	-0.0017608854	728	217	
83	estate	-0.0017540514	2716	784	
84	gain	-0.0015642030	6247	1852	
85	share	-0.0014887732	9697	2432	
86	fixedrate	-0.0013036745	812	303	
87	conditions	-0.0012515516	8856	2589	
88	supreme	-0.0009492661	237	66	
89	longtail	-0.0007928385	8	4	
90	dividend	-0.0007362741	3427	896	
91	improved	-0.0002877444	5524	1315	
92	$	-0.0002395466	6309	2172	
93	authorized	-0.0001037615	2525	761	
94	association	-0.0000118954	1428	368	
95	forbearance	0.0004112238	102	39	
96	warrants	0.0006607219	2643	731	
97	material	0.0006818501	10675	2626	
98	ounce	0.0007999596	64	8	
99	failure	0.0010127209	5127	1160	
100	lenders	0.0013683647	1749	576	
101	covenant	0.0014390453	979	418	
102	experienced	0.0018634726	7046	1848	
103	debenture	0.0019784863	374	91	
104	overhead	0.0020508185	3120	909	
105	travel	0.0023071405	2021	721	
106	loss	0.0027782258	10556	2703	
107	default	0.0032899859	1867	777	
108	systems	0.0041423027	9106	1707	
109	compliance	0.0041780755	6995	1472	
110	concern	0.0045060500	967	312	
111	deficit	0.0045681427	1551	535	
112	iii	0.0046613327	3878	837	
113	annum	0.0046780384	2017	534	
114	raise	0.0047650495	2267	841	
115	software	0.0052244650	7036	1179	
116	oo	0.0057821096	10	1	
117	yarn	0.0058982809	42	10	
118	fabrics	0.0070229885	80	19	
119	issue	0.0070397354	6091	1282	
120	financing	0.0077108523	10132	2538	
121	additional	0.0082449947	11566	2752	
122	problems	0.0083942525	4789	711	
123	company	0.0085525169	11932	2748	
124	raising	0.0108728471	811	289	
125	waiver	0.0125430341	686	239	
126	convertible	0.0149530247	3317	908	
127	change#	0.0154042005	11	4	
128	compliant	0.0156351700	3895	52	
129	assurance	0.0160497186	7495	1714	
130	administrative	0.0161868978	10877	2555	
131	iru	0.0172323025	20	7	
132	negotiating	0.0176258733	927	261	
133	computer	0.0199994197	7106	1037	
134	contemplates	0.0206389257	227	72	
135	defaults	0.0208040257	549	241	
136	covenants	0.0224025681	4560	1334	
137	monoline	0.0283932539	10	0	
138	dtr	0.0300474908	5	0	
139	casebasis	0.0318337184	10	1	
140	going	0.0392456886	896	385	
141	perma	0.0436809467	5	0	
142	correspondents	0.0529492481	50	7	
143	westbrook	0.0574437519	8	4	
144	editda	0.0600640901	5	2	
145	starter	0.0791015118	15	5	
146	heapleach	0.0853457216	5	0	
147	elsinore	0.0951553246	5	0	

```
148           careers  0.1115174484        5       7
149        gainonsale  0.1117425141        6       3
150          resultsa  0.1308973807        5       0
151               ive  0.1373336102        5       6
152         NT_before  0.7810684847    12265    2845
Bias term:
        s0
-0.6897472
******************************************************
*                      2003                          *
******************************************************
hi
final_test       2003
alpha    1
dfmax    5000
year_window      5
refine   TRUE
input_dir        /usr2/brendano/10K/feat3countsel/fc5_relprune/2003.relprune
input_ext        tf.num
nohist   FALSE
outfile          results.RData
final_test       2003
final_train      1998 1999 2000 2001 2002
dev_test         2002
dev_train        1998 1999 2000 2001
                 name          beta train_df test_df
1              dinuba -9.078883e-01        5       1
2              dekalb -6.606667e-01        6       4
3         prestocking -4.358477e-01        5       0
4             vickers -3.545241e-01        5       3
5                 hrn -2.930392e-01        5       1
6               gras -2.795364e-01        7       3
7               opis -2.045160e-01        6       2
8             kubota -1.972183e-01        5       1
9         unpublished -1.954184e-01       15       7
10          wentworth -1.879728e-01        9       1
11               otto -1.866197e-01        5       0
12              swine -1.850276e-01        9       6
13             verity -1.614099e-01       11       6
14        accrediting -1.585159e-01       19       8
15    selfconstructed -1.556867e-01        6       2
16        transmittal -1.531388e-01       14       5
17       weyerhaeuser -1.395552e-01        6       3
18             onpeak -1.381719e-01        6       3
19              murex -1.368265e-01        6       1
20          kingsport -1.290701e-01       12       5
21          airliners -1.143230e-01        9       1
22    recertification -1.067868e-01       12       7
23             raytel -1.058415e-01        5       1
24              zeiss -1.025905e-01       11       2
25    complimentaries -8.134368e-02        8       3
26                hac -7.226874e-02        5       0
27           validly -7.039252e-02       39      17
28          urologist -6.357133e-02       13       2
29           nonaudit -6.009197e-02       12      32
30         cofounders -5.459214e-02       24       9
31               cgac -5.427497e-02        5       1
32         waterfront -4.994192e-02       15       6
33    selfsufficiency -4.749302e-02        6       4
34             warmer -4.187371e-02      209      63
35       samelocation -4.064180e-02        6       1
36               tuna -3.750493e-02        5       2
37            rewrite -3.716120e-02       20       3
38           parkside -3.425454e-02        8       2
39           rockaway -3.390767e-02        5       4
40  interstatejohnson -3.106873e-02        5       1
41         rescission -3.033574e-02      304    1179
```

42	dkny	-3.012195e-02	9	4
43	surviving	-3.005426e-02	265	74
44	qualitative	-2.969947e-02	9775	3403
45	corrections	-2.570325e-02	444	1077
46	karan	-2.316005e-02	9	2
47	scholarships	-2.261725e-02	7	1
48	hawesville	-2.170804e-02	5	1
49	rates	-1.583691e-02	11334	3513
50	unanimously	-1.559059e-02	89	47
51	rate	-1.550357e-02	12110	3523
52	ultraseek	-1.550183e-02	5	1
53	netinterest	-1.505631e-02	7	3
54	risk	-1.349971e-02	11607	3559
55	enact	-1.315285e-02	89	55
56	reits	-1.298526e-02	317	85
57	repurchases	-1.276083e-02	1578	713
58	properties	-1.252202e-02	3423	1284
59	ffo	-1.177811e-02	311	82
60	reit	-1.102034e-02	516	162
61	doctrines	-1.053009e-02	7	4
62	adopted	-1.036130e-02	5220	2644
63	tender	-9.151194e-03	1090	344
64	airfares	-9.127491e-03	6	5
65	weather	-9.096669e-03	1924	654
66	income	-7.839979e-03	12707	3571
67	merger	-7.802669e-03	3481	1001
68	estate	-7.788409e-03	3043	1252
69	market	-7.357373e-03	12434	3601
70	improved	-5.519302e-03	5759	1895
71	tenants	-5.389849e-03	488	181
72	conditions	-4.733812e-03	9971	3394
73	unitholders	-4.605925e-03	173	63
74	strong	-4.208183e-03	3827	1289
75	distributions	-3.892319e-03	2010	684
76	fullprice	-3.859052e-03	16	8
77	contents	-3.171649e-03	586	1071
78	board	-3.016645e-03	8771	3091
79	gains	-2.706569e-03	6096	2396
80	splitrock	-2.655944e-03	5	1
81	improvements	-2.478335e-03	5691	1906
82	repayments	-2.324962e-03	2327	881
83	property	-2.196424e-03	8298	2904
84	share	-1.811387e-03	10347	3204
85	propane	-1.804960e-03	130	42
86	changes	-1.797283e-03	11549	3517
87	nareit	-1.714156e-03	276	69
88	earnings	-1.637201e-03	9545	2985
89	#table	-1.606426e-03	515	1049
90	gain	-1.026585e-03	7091	2495
91	units	-7.942758e-04	3945	1728
92	excluding	-3.367508e-04	6338	2131
93	fair	-3.338553e-05	8475	3375
94	expects	6.939031e-05	7825	1783
95	working	5.335098e-04	10113	2863
96	fabrics	6.567963e-04	82	28
97	wages	6.913030e-04	1193	426
98	travel	9.731503e-04	2436	944
99	divisions	1.223674e-03	1086	353
100	warrants	1.358643e-03	2935	911
101	raising	1.975282e-03	972	364
102	software	2.122688e-03	7626	1441
103	compliance	2.548305e-03	7777	2113
104	negotiating	2.785688e-03	1012	333
105	page	2.857022e-03	10630	1660
106	problems	3.025478e-03	5188	985
107	plums	3.115779e-03	6	0
108	vendors	3.380067e-03	4750	936

```
109           company  3.595986e-03    12464    3499
110          resultsa  3.804925e-03        5       1
111     noncompliance  3.943450e-03     1277     288
112         debenture  4.412836e-03      390     104
113              loss  4.691773e-03    11416    3495
114           systems  5.175704e-03     9766    2116
115            minute  6.421213e-03      175      63
116             issue  6.666588e-03     6890    2301
117        additional  6.806909e-03    12232    3537
118           lenders  6.865557e-03     2063     798
119             annum  8.075114e-03     2217     725
120           default  8.274151e-03     2391    1079
121    administrative  9.149284e-03    11441    3220
122         financing  1.049313e-02    10921    3225
123       unqualified  1.100722e-02       14      10
124           openpit  1.128016e-02        8       2
125             raise  1.360790e-02     2819    1118
126       forbearance  1.627533e-02      129      48
127          covenant  1.646586e-02     1254     592
128              yarn  1.741604e-02       44      12
129         covenants  1.747949e-02     5134    1843
130         compliant  1.764452e-02     3895      67
131           concern  1.783118e-02     1171     437
132          computer  1.837160e-02     7485    1230
133             doubt  1.926797e-02      473     199
134            waiver  1.980139e-02      843     325
135       convertible  2.227315e-02     3663    1145
136         assurance  2.456835e-02     8023    2228
137        exhibitors  2.650767e-02        9       6
138           change#  2.683650e-02       14       7
139               iru  2.872490e-02       27       7
140             perma  3.061636e-02        5       0
141          jacobson  4.131675e-02       11       2
142             going  4.411245e-02     1170     500
143         casebasis  5.542761e-02       11       0
144          backhaul  7.383069e-02       16       9
145                #n  8.438790e-02       11       7
146    correspondents  1.292204e-01       47      11
147            editda  5.134741e-01        7       0
148         NT_before  7.803358e-01    12850    3611
Bias term:
        s0
-0.6409522
******************************************************
*                        2004                        *
******************************************************
hi
final_test       2004
alpha    1
dfmax    5000
year_window      5
refine   TRUE
input_dir        /usr2/brendano/10K/feat3countsel/fc5_relprune/2004.relprune
input_ext        tf.num
nohist   FALSE
outfile          results.RData
final_test       2004
final_train      1999 2000 2001 2002 2003
dev_test         2003
dev_train        1999 2000 2001 2002
                   name          beta train_df test_df
1                 dinuba -0.7394264800        6       1
2                   coca -0.4339536509        5       2
3              wentworth -0.3931901212        8       2
4                clothes -0.3749339596        6       2
5                vickers -0.3410063723        7       3
6                  balsa -0.3005583284        5       0
```

7	raytel	-0.2889556905	5	1	
8	hrn	-0.2843768536	6	0	
9	autoinjectors	-0.2150419406	5	1	
10	bonusing	-0.2128270396	9	3	
11	digester	-0.1999214840	5	1	
12	parkside	-0.1952768307	8	0	
13	westside	-0.1633314733	10	4	
14	scios	-0.1406355509	11	1	
15	rentfree	-0.1268255498	8	1	
16	sithe	-0.1156543170	5	0	
17	karan	-0.1148782572	9	1	
18	goingprivate	-0.1100508594	9	4	
19	mlmci	-0.1062797945	5	0	
20	selfconstructed	-0.1032177441	6	3	
21	talc	-0.0994808935	9	4	
22	ebeam	-0.0909897133	6	5	
23	airliners	-0.0858988067	7	0	
24	scholarships	-0.0855632228	7	4	
25	tuna	-0.0820060021	6	2	
26	onpeak	-0.0802205302	9	7	
27	weyerhaeuser	-0.0782696082	7	7	
28	unpublished	-0.0775022111	20	9	
29	unanimously	-0.0744315821	117	43	
30	travelocity	-0.0674244217	7	2	
31	algonquin	-0.0653734775	11	3	
32	trolley	-0.0640759661	8	2	
33	accrediting	-0.0583874513	23	6	
34	nonmilitary	-0.0570685890	14	7	
35	liquefied	-0.0536685400	40	32	
36	transmittal	-0.0523825688	18	5	
37	statementprospectus	-0.0467863630	34	7	
38	verity	-0.0369332447	14	2	
39	saver	-0.0365996877	5	3	
40	posting	-0.0350328215	128	64	
41	zeiss	-0.0339244033	13	3	
42	transgas	-0.0321504472	5	0	
43	bottles	-0.0296998471	73	23	
44	inactivity	-0.0273516043	6	6	
45	opis	-0.0259839803	8	2	
46	rewrite	-0.0256227186	21	4	
47	archway	-0.0253113329	5	0	
48	upgradesenhancements	-0.0252077297	7	2	
49	antennae	-0.0247885050	29	9	
50	surviving	-0.0247232966	290	76	
51	brooding	-0.0242917286	7	1	
52	urologist	-0.0222328574	13	1	
53	kingsport	-0.0201696568	16	5	
54	molson	-0.0196823067	5	3	
55	pension	-0.0168075169	1518	908	
56	aws	-0.0166408556	5	0	
57	yankee	-0.0160064881	32	13	
58	fin	-0.0158213160	1834	1763	
59	properties	-0.0156217205	4089	1398	
60	hawkeye	-0.0143234634	6	4	
61	iraq	-0.0140439820	269	284	
62	exit	-0.0136387541	3646	1161	
63	fastforward	-0.0134183610	5	2	
64	williams	-0.0133651083	120	49	
65	proposed	-0.0116325476	2791	1098	
66	adopted	-0.0115567596	7128	2346	
67	stockbased	-0.0114740294	2772	1155	
68	rates	-0.0109616059	13069	3470	
69	validly	-0.0107847551	53	15	
70	gains	-0.0101118228	7505	2279	
71	repurchases	-0.0099361141	2086	781	
72	qtr	-0.0096784345	54	18	
73	disclosures	-0.0096301033	11462	3239	

74	interpretation	-0.0092839936	3151	2132
75	policies	-0.0091836000	8782	3449
76	income	-0.0089767883	13848	3517
77	tenants	-0.0085080350	597	207
78	assumptions	-0.0083367897	7963	3231
79	warmer	-0.0080538734	241	67
80	merger	-0.0079606364	3905	939
81	ellis	-0.0078920701	44	9
82	ffo	-0.0076187495	340	84
83	performancebased	-0.0072258796	299	126
84	reconciliation	-0.0070318753	886	513
85	weather	-0.0069319946	2224	744
86	contents	-0.0068658912	1648	1378
87	deadlines	-0.0065132421	132	46
88	tension	-0.0064120896	32	18
89	interim	-0.0058508906	3591	1684
90	recertification	-0.0057057610	19	5
91	shelf	-0.0056619374	1139	507
92	rate	-0.0056396830	13461	3484
93	audits	-0.0056247174	751	445
94	$	-0.0056182060	9571	3400
95	war	-0.0055435030	873	604
96	%	-0.0048624820	4997	2335
97	liability	-0.0047271435	8642	3032
98	transported	-0.0045514136	229	109
99	disclosure	-0.0045208399	6861	2545
100	conditions	-0.0039741113	11672	3394
101	contributed	-0.0038775020	5542	1795
102	actuarial	-0.0038256134	879	680
103	utility	-0.0037202498	1581	484
104	incentive	-0.0035761247	2917	1136
105	bcf	-0.0034730954	175	71
106	returns	-0.0032526027	4076	1979
107	improved	-0.0031607146	6496	2191
108	estimation	-0.0030856843	1074	675
109	driven	-0.0029433127	2598	1246
110	circumstances	-0.0029282787	7266	3027
111	estimates	-0.0026550688	9707	3410
112	fructose	-0.0026088750	10	4
113	when	-0.0025525964	11349	3421
114	repayments	-0.0024481144	2844	922
115	probable	-0.0024476055	3123	1547
116	types	-0.0024263292	4083	1695
117	oneforsix	-0.0024123121	5	1
118	earnings	-0.0023142517	10814	3055
119	ratings	-0.0022712220	1485	686
120	excluding	-0.0022402110	7379	2159
121	regulated	-0.0021109448	1170	483
122	guarantees	-0.0020498640	3590	1569
123	estate	-0.0020404619	3795	1315
124	eitf	-0.0019432182	2391	1258
125	audit	-0.0019383174	2043	1188
126	rescission	-0.0019004062	1471	444
127	summary	-0.0016730453	3749	1738
128	insurance	-0.0016015722	7376	2691
129	generally	-0.0015338474	11665	3436
130	regulatory	-0.0012744528	6165	2020
131	contractual	-0.0012621872	6782	3315
132	corrections	-0.0011633232	1481	384
133	counterparties	-0.0011374147	852	351
134	causal	-0.0010151019	5	1
135	units	-0.0009817621	4980	1857
136	unit	-0.0009663453	5614	2010
137	recover	-0.0009251756	2287	1028
138	increases	-0.0008783999	11751	3185
139	strong	-0.0008123598	4462	1625
140	considers	-0.0007921584	2708	1020

141	billion	-0.0007036017	2710	1004	
142	index	-0.0005805642	1265	512	
143	judgments	-0.0005734867	4357	2649	
144	both	-0.0005609407	11732	3351	
145	political	-0.0004606900	3552	1273	
146	gas	-0.0004333151	1631	584	
147	calculated	-0.0002758988	3644	1514	
148	paper	-0.0001554779	2633	792	
149	rating	-0.0001391009	1788	772	
150	#year	-0.0001039526	1528	630	
151	startup	0.0002935420	2582	520	
152	countdown	0.0003520008	8	2	
153	addressed	0.0005508903	1454	319	
154	issues	0.0008718181	8561	2281	
155	will	0.0016885866	13898	3545	
156	default	0.0017684282	3168	1132	
157	covenant	0.0017747603	1680	601	
158	combinations	0.0021381580	3617	797	
159	supersedes	0.0021716043	1599	210	
160	ziffdavis	0.0026264979	5	1	
161	additional	0.0029001459	13441	3507	
162	systems	0.0032946974	9760	2204	
163	warrants	0.0033899881	3341	963	
164	usdata	0.0034715854	5	0	
165	covenants	0.0040421756	6093	1874	
166	working	0.0041142979	11049	2877	
167	placements	0.0042460255	1049	272	
168	date	0.0047380876	12343	3332	
169	lenders	0.0048052761	2550	800	
170	debenture	0.0048886305	419	142	
171	minute	0.0050566317	210	79	
172	vendors	0.0051205932	4810	1046	
173	apart	0.0055713625	624	151	
174	listing	0.0068521240	955	347	
175	loss	0.0070128877	12876	3439	
176	nondebtor	0.0076614877	5	2	
177	page	0.0078088770	10085	1321	
178	distance	0.0090887132	576	149	
179	annum	0.0095229138	2538	733	
180	rhodes	0.0095402228	18	4	
181	consoli	0.0099451442	8	2	
182	administrative	0.0119798401	12488	3177	
183	computer	0.0127436595	6880	1228	
184	molex	0.0130939468	9	3	
185	readiness	0.0138744474	2122	37	
186	smallcap	0.0142453390	338	105	
187	financing	0.0151527437	12187	3248	
188	assurance	0.0152654832	8694	2220	
189	nitric	0.0155444921	5	2	
190	problems	0.0158096674	5228	1048	
191	waiver	0.0168824423	1046	318	
192	noncompliance	0.0188884477	1382	324	
193	raise	0.0194501000	3570	1168	
194	forbearance	0.0218853664	162	31	
195	delisting	0.0224769359	371	98	
196	iru	0.0240809000	31	10	
197	doubt	0.0244981005	615	179	
198	unqualified	0.0248990648	23	25	
199	dudley	0.0269496249	6	3	
200	correspondents	0.0277652776	48	8	
201	compliant	0.0309479075	2816	106	
202	going	0.0312943092	1549	538	
203	concern	0.0342715332	1461	411	
204	imaged	0.0367368485	9	6	
205	eightyfour	0.0387363847	6	3	
206	creditimpaired	0.0408268452	11	3	
207	accor	0.0456357080	7	3	

```
208               azurix  0.0570607408        5       1
209              solutia  0.0775414422        9       3
210               weakly  0.0778798940        7       2
211      pennsylvanianew  0.1124502508        7       0
212           colocating  0.1189506919        9       2
213             backhaul  0.1201602882       25      14
214                plums  0.1646265316        5       0
215                 kets  0.1747366593        6       0
216            casebasis  0.2727789445       10       1
217               editda  0.7072108249        6       1
218            NT_before  0.8052296688    14000    3558
Bias term:
        s0
-0.6473796
******************************************************
*                      2005                          *
******************************************************
hi
final_test       2005
alpha    1
dfmax    5000
year_window      5
refine   TRUE
input_dir        /usr2/brendano/10K/feat3countsel/fc5_relprune/2005.relprune
input_ext        tf.num
nohist   FALSE
outfile          results.RData
final_test       2005
final_train      2000 2001 2002 2003 2004
dev_test         2004
dev_train        2000 2001 2002 2003
          name          beta train_df test_df
1     corvette -0.0531057955        7       0
2  assumptions -0.0140119153    10239    3177
3         exit -0.0089093164     4623     711
4        rates -0.0062705532    14328    3397
5     policies -0.0059739000    11392    3355
6            $ -0.0059505800    11629    3337
7    estimates -0.0055229472    11521    3334
8   guarantees -0.0054450764     4892    1182
9   stockbased -0.0050611227     3884    1884
10   liability -0.0015393517    10522    2935
11  disclosure -0.0010818314     8455    2495
12      change -0.0009642059    13593    3321
13           % -0.0008966848     6766    2494
14     concern  0.0063469374     1624     427
15   NT_before  0.8816442182    15034    3474
Bias term:
        s0
-0.4246537
******************************************************
*                      2006                          *
******************************************************
hi
final_test       2006
alpha    1
dfmax    5000
year_window      5
refine   TRUE
input_dir        /usr2/brendano/10K/feat3countsel/fc5_relprune/2006.relprune
input_ext        tf.num
nohist   FALSE
outfile          results.RData
final_test       2006
final_train      2001 2002 2003 2004 2005
dev_test         2005
dev_train        2001 2002 2003 2004
```

```
              name          beta train_df test_df
1         corvette -0.5144980899        6       1
2           dinuba -0.2554072200        6       1
3           syncra -0.0821441548        5       0
4  hartscottrodino -0.0468193960      100      23
5           havent -0.0266532942        7       5
6           merger -0.0216250665     4467     824
7          envenue -0.0172499818        5       0
8           ruston -0.0158597404        5       0
9              sub -0.0153235152      224      59
10           rates -0.0113903288    15541    3231
11            exit -0.0088433495     5109     605
12           gains -0.0063443036     9739    2131
13       guarantor -0.0060000221     3118     263
14         billion -0.0056219579     3952    1073
15          income -0.0038738462    15912    3277
16      guarantees -0.0037436990     5796    1088
17            rate -0.0036034426    15652    3234
18      disclosure -0.0032723212    10215    2278
19         ratings -0.0032719731     2547     725
20       surviving -0.0027355549      350      73
21        earnings -0.0023550347    13348    2898
22        proposed -0.0023185943     4188     883
23               % -0.0022305705     8631    2445
24     switchboard -0.0022059460        7       0
25         changes -0.0002398087    15516    3240
26           doubt  0.0044972412      811     156
27           going  0.0047095913     2261     524
28          editda  0.0083087818        8       2
29    combinations  0.0119955395     4852     621
30           raise  0.0121935776     5024    1035
31        delisted  0.0134963363      561      44
32         concern  0.0401072562     1816     347
33       NT_before  0.8331310861    16084    3306
Bias term:
        s0
-0.6181522
```