Credibility and Risk Classification

Casualty Actuarial Society / Canadian Institute of Actuaries

Montreal

November 17, 2004

Presented by Christopher Monsour, FCAS, MAAA

Goals

- Establish a connection between three things:
  - Credibility
  - Penalized regression
  - Random effects models
- Show how the cross-validation idea for penalized regression can be applied to credibility to estimate K

Classic example

Auto Insurance

- Credibility weight territorial or type of car experience with experience for entire state
- Loss Ratios or pure premiums or adjusted pure premiums
- Two major questions:
  - How to determine the optimal number of classes / class boundaries?
  - How to determine the credibility to give each class?
    - Or, more simply, assuming one is going to use the formula N/(N+K), how to determine K?
    - It's the ratio of within variance to between variance, but is estimating the variances the right way to estimate K?
      - Related question: Are you measuring N appropriately?
  - Note that generalized linear models as usually used do not provide all the answers

Another way to view credibility

Random Effects Model

- Just like a regression model, but there is a prior belief for the parameters
- Consider the simplest case: Two classes
  - Let x_i = 0 or 1, indicating the class of policy i
  - Let y_i be the corresponding pure premium or loss ratio
  - Center X and Y by subtracting the mean from each. (This eliminates the need for a constant term in the regression). For example, now each x_i is equal to $-\pi$ or $1-\pi$ where π is the proportion of class 1
- Then the regression model is $Y=\beta X$, and the negative log-likelihood is equal to a constant plus the following expression:
  - $\Sigma_i(y_i-\beta x_i)^2/\sigma^2$, which is proportional to $\Sigma_i(y_i-\beta x_i)^2$
- If β is taken as $N(0,\tau^2)$ and X is $N(0,\sigma^2)$ (i.e., τ^2 is between variance and σ^2 is within variance), the negative log-likelihood is then
  - $\Sigma_i(y_i-\beta x_i)^2/2\sigma^2 + \beta^2/2\tau^2$, which is proportional to $\Sigma_i(y_i-\beta x_i)^2 + (\sigma^2/\tau^2)\beta^2$

Another way to view credibility

Ridge Regression

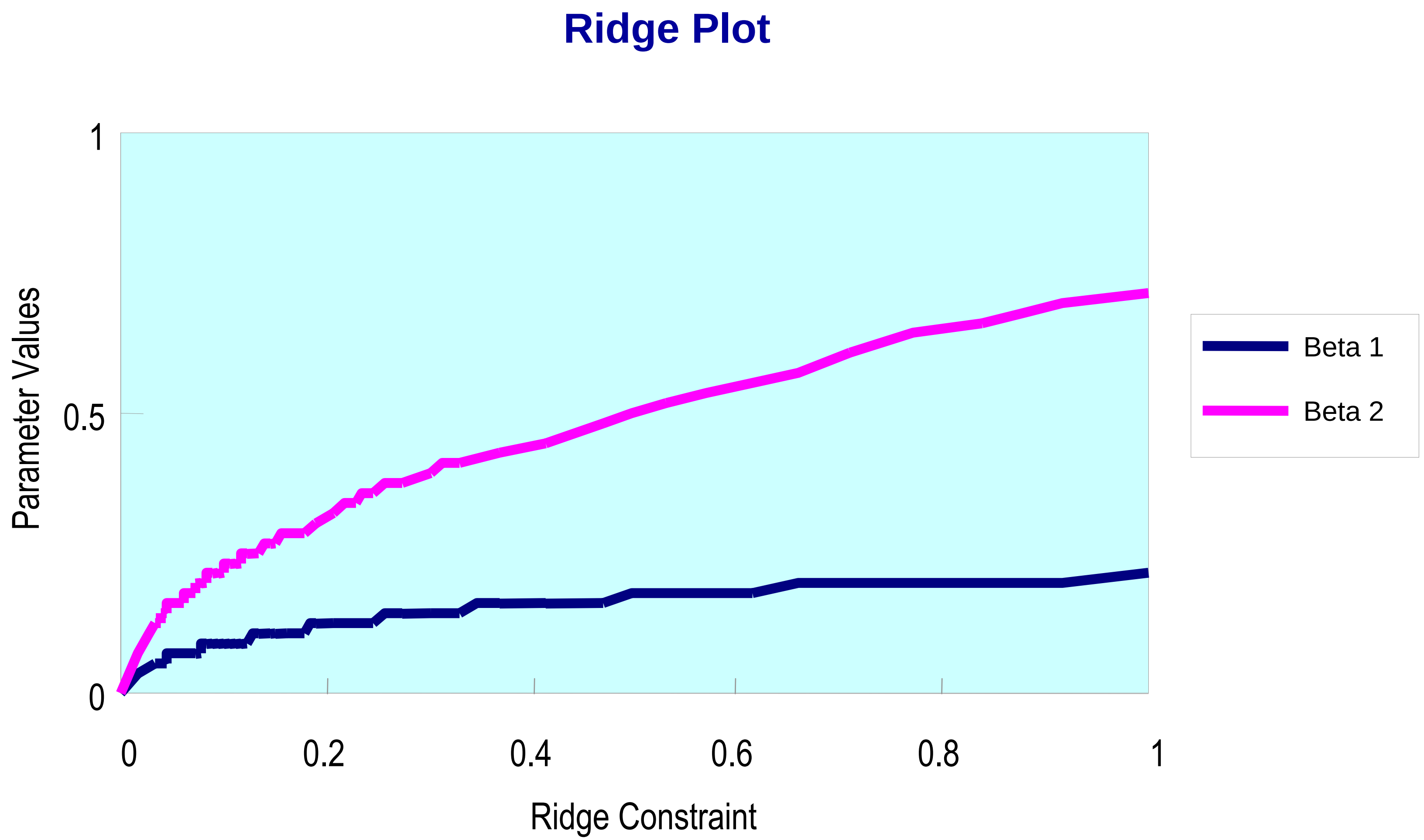
- Used to penalize large parameters, using sum of squares of parameter sizes as the penalty
- Center Y
- Center **and standardize each** X_i (divide by standard deviations), separately for each i
- Equation to minimize is
  - $\Sigma_i(y_i-\Sigma_j\beta_jx_{ij})^2$ subject to $\Sigma_j\beta_j^2 < \Lambda$
    — Equivalent to minimizing $\Sigma_i(y_i-\Sigma_j\beta_jx_{ij})^2 + \lambda\Sigma_j\beta_j^2$ with $\lambda > 0$
- Goal of ridge regression is to control for multicollinearity
  - This is the reason for standardizing the predictor variables
  - In credibility applications, one does not want to standardize the predictors, as they are class indicators

Another way to view credibility

Ridge Regression and Credibility

- For one variable, the solution for β is the ordinary least squares solution times
  - $(\Sigma_i x_i^2) / \{(\sigma^2/\tau^2) + \Sigma_i x_i^2\}$
  - Corresponds to $M\pi(1-\pi) / \{ (\sigma^2/\tau^2) + M\pi(1-\pi) \}$, where M is the total number of observations
  - When the total number is much larger than the number $N=M\pi$ in the rare class, this reduces to N / (K+N), exactly the Bayesian credibility result
  - Note that this doesn't mean classical credibility is wrong for large classes…just that the variance of β no longer precisely corresponds to the between variance in that case
  - Of course, we knew this would happen because normal-normal is a conjugate pair

Another way to view credibility

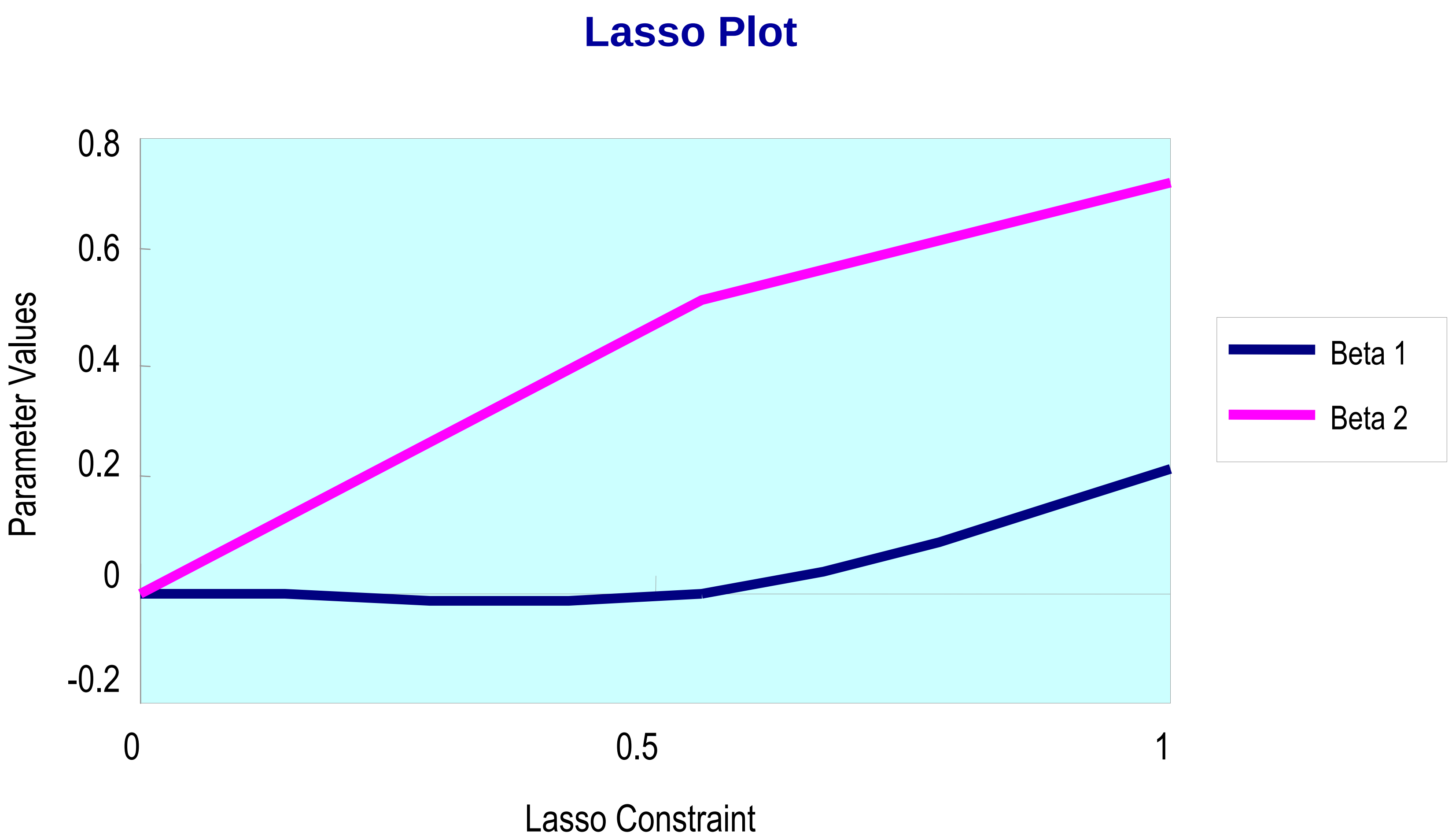


Another way to view credibility

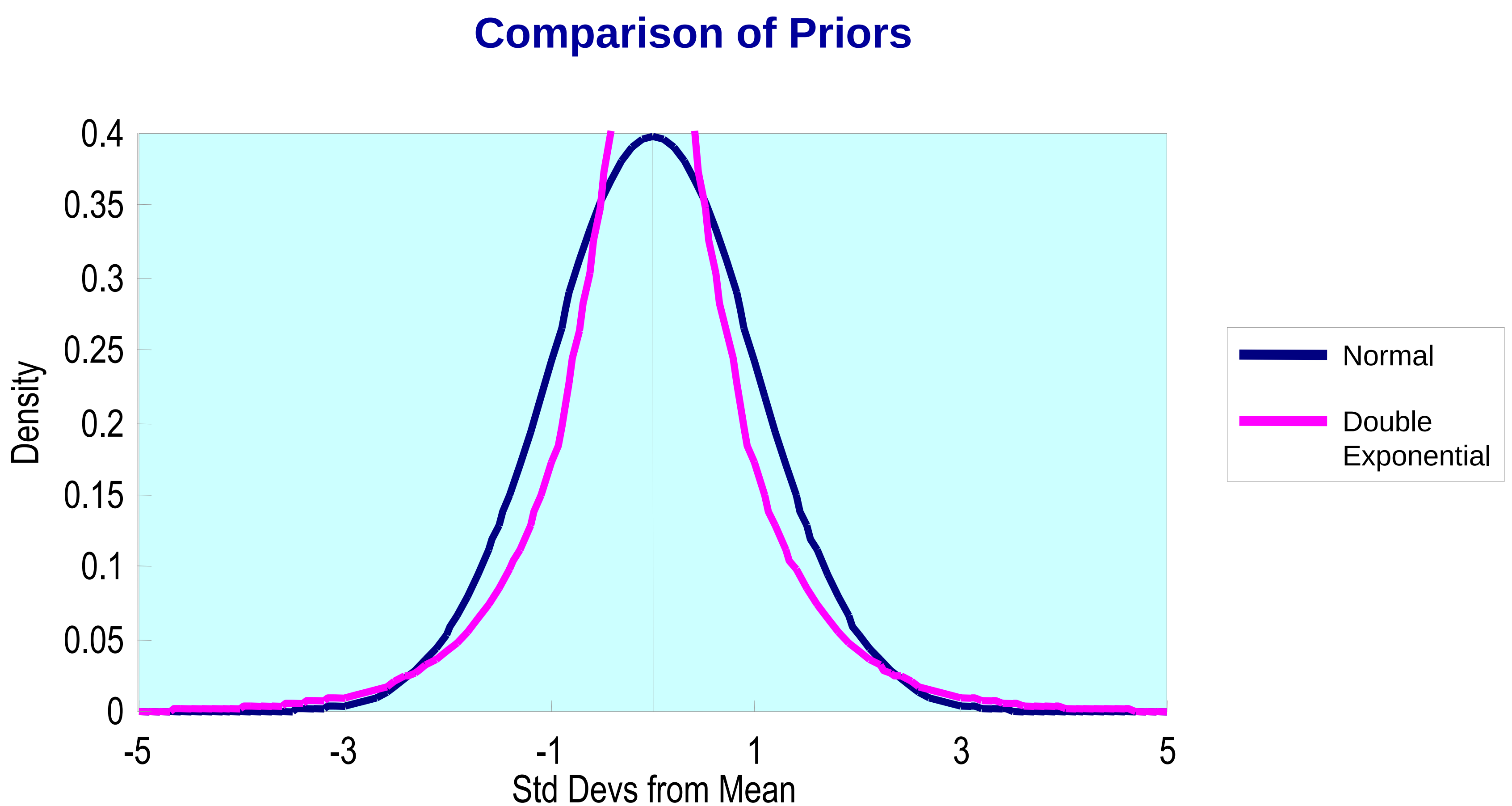
The Lasso

- Unlike ridge, can shrink some parameters all the way to 0
- Penalty is sum of absolute parameter values, i.e., minimize
  - $\Sigma_i(y_i-\Sigma_j\beta_j x_{ij})^2$ subject to a constraint $\Sigma_j|\beta_j|<\Lambda$
  - This corresponds to minimizing $\Sigma_i(y_i-\Sigma_j\beta_j x_{ij})^2 + \lambda\Sigma_j|\beta_j|$ with $\lambda>0$
- In Bayesian interpretation, corresponds to prior for each β_i that is double exponential with a density of $(\sigma^2/\lambda)\exp(-\lambda|\beta_i|/2\sigma^2)$
  - Note that $\text{var}(\beta_i)=4\sigma^2/\lambda$. Call this τ^2
  - This corresponds to a more diffuse (more tail-heavy) prior than the normal
- Again, in "standard" penalized regression, one centers Y and centers and standardizes each X_i. In credibility-type applications, one would only center, but not standardize
- Equivalent to moving every parameter estimate toward zero by 2K/N, where K is σ^2/τ^2 and N is the size of the class. Large parameter values are drawn back toward the mean less than they would be by a credibility multiple

Another way to view credibility



Another way to view credibility



Another way to view credibility

Comparison of “Credibility” Methods

Unbiased estimate of beta = 100

		Normal	Dbl Exponential
		Prior	Prior
N	K	Z	Z
10	2000	0.50	0.00
20	2000	0.99	50.00
50	2000	2.44	80.00
100	2000	4.76	90.00
200	2000	9.09	95.00
500	2000	20.00	98.00
1000	2000	33.33	99.00

Unbiased estimate of beta = 10

		Normal	Dbl Exponential
		Prior	Prior
N	K	Z	Z
10	2000	0.50	0.00
20	2000	0.99	0.00
50	2000	2.44	0.00
100	2000	4.76	0.00
200	2000	9.09	50.00
500	2000	20.00	80.00
1000	2000	33.33	90.00

What's wrong with a lasso (double exponential) version of credibility?

- It does exhibit strange behavior, as you can see above
- It depends on the scaling of the Y variable. Essentially this means that you have to test different K values and determine the one you like
- Maximum likelihood estimates give posterior **mode**, not posterior **mean**
  - For normal, these are the same
- On the other hand, the lasso can be useful as a selection method, for determining what distinctions to include, since it does force some parameters to zero
- Equivalent to ranking the classes by βN, i.e., by their difference from the global mean times their size, and choosing a threshold on this ranked list for including a given class as being substantially different from the global mean
  - If you can order the classes, can use a similar condition for adjacent classes

Cross-validation

- λ in penalized regression traditionally determined by cross-validation
- Divide data into N pieces at random
  - For each piece, estimate model from the other N-1 pieces, and test its fit (e.g., sum of squared errors) on the piece
  - Add up these sum of square errors
  - Plot vs. λ
- Can do the same thing with K in credibility
- Can even simply use goal-seek to minimize out-of-sample (cross-validation) squared error and minimize K
- Minimizing squared error on in-sample data results in overfit
  - For example, testing on in-sample data always results in preferring two classes to one

Cross-validation example

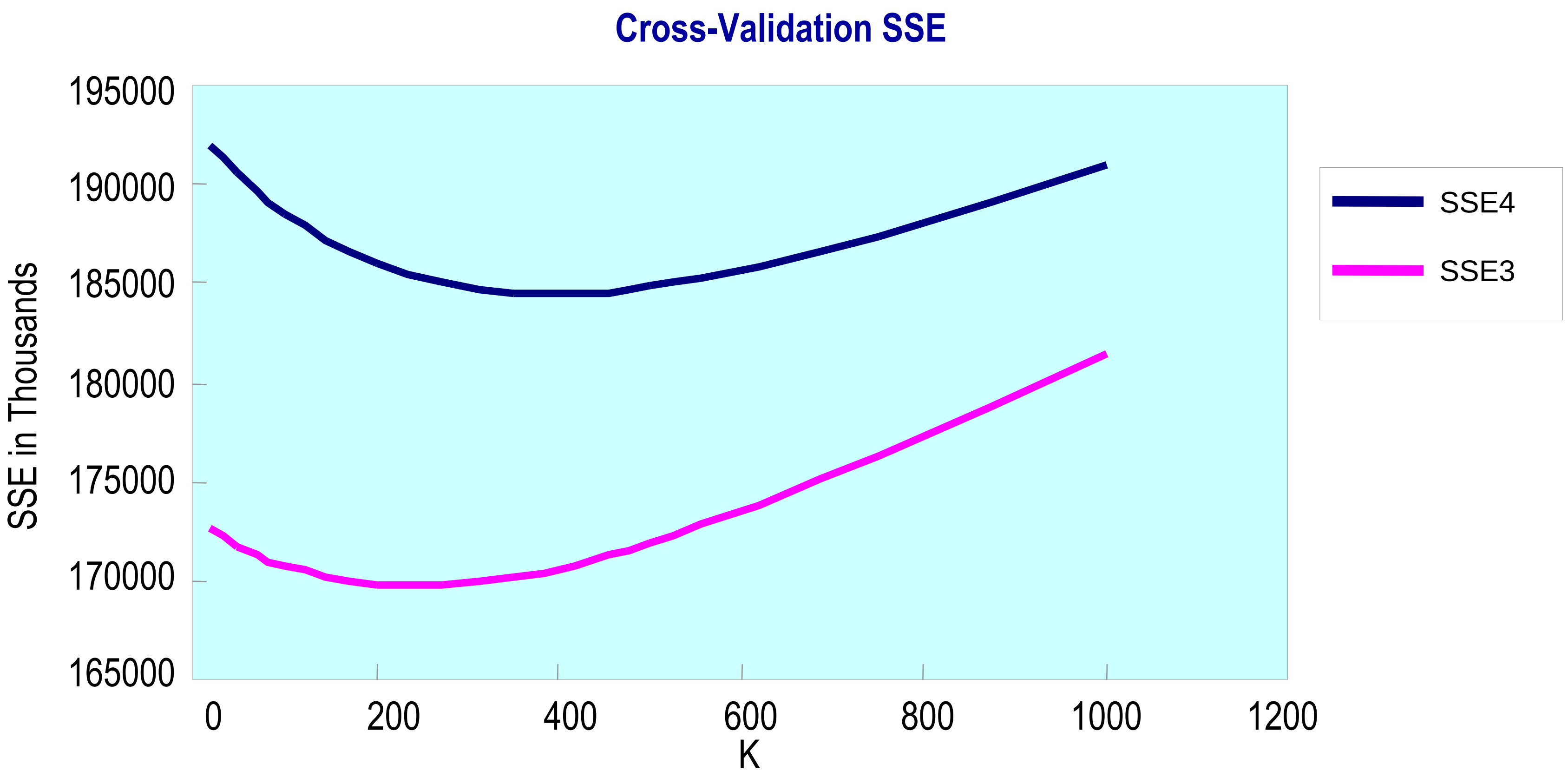
Origin of Data in Spreadsheet

- Auto Insurance
- Based on Turner's paper in the 2004 Ratemaking Forum
- Created random claims data to duplicate the means and variances in his data
- Then divided this into cross-validation folds
  - See Exhibit on last page of handout

- Can try combining various combinations of classes into one and compare the cross-validation SSEs.
- It is possible for the best fit to come from no divisions at all
  - In this case, combining 2 and 3 is best (the "three class" example in the slide below)

Cross-validation example

- Chart of cross-validation SSE versus K for four and three class examples
- K's estimated by Turner from the actual within cell variances are 675 for the four class example and 504 for the three class example



Cross-validation and credibility

- Cross-validation is a reasonable method to estimate any “tuning” parameter
- It often makes sense to us separate years of data as the “folds” in the cross-validation, since it’s predictive value across years that matters most
- Credibility is no longer a unique animal

  As a profession we need to become acquainted with its cousins

Cross-Validation
for credibility
Four class Example

VARY K 395 TO MINIMIZE Cross-validation SSE (sum of col (8)) 184490992

Risk Class	Cross-Validation Fold	Exposures** (1)	Average Loss (2)	Total Loss** (3)	Other folds Class Mean (4)	Other folds Global Mean (5)	Z* N/(N+K) (6)	Credibility Estimate (7)	(Grouped) SSE (8) (1) x ((7)-(2))^2
1	1	362	749.75	271,411	721.06	937.08	0.777	769.16	136348
	2	354	606.93	214,852	757.68	897.64	0.778	788.70	11697027
	3	354	586.88	207,756	762.80	950.92	0.778	804.50	16764109
	4	328	751.26	246,413	721.40	940.70	0.782	769.31	106892
	5	343	948.46	325,322	672.70	929.73	0.780	729.32	16471265
2	1	277	928.60	257,223	915.66	937.08	0.758	920.85	16658
	2	284	832.15	236,331	937.86	897.64	0.757	928.08	2613674
	3	323	972.35	314,069	903.30	950.92	0.751	915.16	1056531
	4	332	1114.19	369,911	862.93	940.70	0.750	882.41	17835616
	5	298	712.63	212,363	968.37	929.73	0.755	958.89	18072666
3	1	294	890.17	261,710	944.69	937.08	0.746	942.76	813038
	2	272	1150.89	313,042	883.78	897.64	0.750	887.25	18906339
	3	316	873.89	276,148	950.25	950.92	0.743	950.42	1851130
	4	294	838.76	246,597	957.69	940.70	0.746	953.38	3862285
	5	280	935.49	261,938	933.25	929.73	0.749	932.36	2746
4	1	298	1131.88	337,300	1150.84	937.08	0.768	1101.35	277823
	2	310	1461.50	453,065	1072.35	897.64	0.767	1031.61	57288986
	3	315	1120.98	353,109	1153.74	950.92	0.766	1106.31	67813
	4	332	1127.74	374,410	1152.42	940.70	0.764	1102.40	213162
	5	354	927.02	328,165	1209.47	929.73	0.761	1142.50	16436886
Total All Classes	1	1248	907.16	1,132,131	937.08				
	2	1208	1071.19	1,294,002	897.64				
	3	1301	855.62	1,113,160	950.92				
	4	1248	892.64	1,114,016	940.70				
	5	1257	936.65	1,177,364	929.73				

* Value of N for credibility formula is based on the number of observations in the other folds
** Data in total for each risk class based on data in Turner's 2004 Ratemaking Forum paper "Credible Risk Classification"
Division into cross-validation folds is illustrative