Forecast Combination

- In the press, you will hear about "Blue Chip Average Forecast" and "Consensus Forecast"
- These are the averages of the forecasts of distinct professional forecasters.
- Is there merit to averaging (combining) different forecasts?
- Or is it better to focus on selecting the best forecast?

GDP Forecast

- Let's consider forecasting GDP growth for 2010Q1 based on data up through 2009
- GDP growth for the four quarters of 2009

2009Q1	2009Q2	2009Q3	2009Q4
-6.4%	-0.7%	2.2%	5.6%

Models

- In p.s. #9, you will estimate models for GDP
  - AR(3) plus 3 lags of *dt3*
  - AR(3) plus 3 lags of *dt12*
  - AR(3) plus 3 lags of *spread12*
  - AR(3) plus 3 lags of *spread120*
  - AR(3) plus 3 lags of *junk*
- Let's reconsider the number of lags
  - (Note: The estimates which follow use a different sample period than your problem set)

AIC for different lag structures

	junk	yield	lags	
	0	1	2	3
AR(1)	571	570	552	554
AR(2)	571	571	552*	554
AR(3)	571	570	552	554

- The model with 2 AR lags and 2 lags of *junk* has the lowest AIC
- But the models with 1 and 3 AR lags have nearly the same AIC
- And the models with 3 lags of *junk* are quite close too

Forecasts

	junk	yield	lags	
	0	1	2	3
AR(1)	4.0	3.8	5.2	4.4
AR(2)	3.9	3.7	5.1*	4.3
AR(3)	4.2	4.1	5.3	4.4

- The point forecasts are quite different
- The model selected by AIC is much higher than the AR model
- The model with 3 lags of *junk* have quite different forecasts

Average Forecast

- The average of the 12 forecasts is

$$\hat{y}_{average} = \frac{4.0+3.9+4.2+3.8+3.7+4.1+5.2+5.1+5.3+4.7+4.3+4.4}{12}$$
$$= 4.4$$

- This is similar to a consensus or Blue Chip forecast.
- You could imagine these 12 forecasts as coming from different forecasters.
- Is it useful to combine the forecasts?

Pseudo Out-of-Sample Experiment

- Split the sample
  - Estimation period: 1954Q2-1994Q4 (30 years)
  - Evaluation period: 1995Q1-2009Q4 (15 years)
- Estimate the 12 models using 1954Q2-1994Q4
  - Fix the parameter estimates
- Use these models to forecast 1995Q1-2009Q4
- Also, take the average forecast for each period
- Create out-of-sample errors for the 12 models
- And the out-of-sample error for the average forecast
- Compare the performance of the methods by RMSE
  - A simplified version of predictive least square (PLS)

Out-of-Sample RMSE

RMSE	junk	yield	lags	
	0	1	2	3
AR(1)	2.46	2.38	2.34	2.34
AR(2)	2.46	2.37	**2.32***	2.32
AR(3)	2.41	2.33	2.36	2.37

RMSE	Average forecast
	2.18

- The comparisons based on out-of-sample RMSE are similar to AIC on full sample
- The lowest RMSE is **2.32**, achieved by the model with 2 lags of each
- But the RMSE of the average forecasts (the average across all 12 forecasts) is **2.18**
- We achieve a much lower RMSE by this simple averaging!
- Why?
- Why is it useful to combine forecasts?
- Can we do better than a simple equal-weighted average?

Theory of Forecast Combination

- Suppose you have forecasts f_1 and f_2 for y
- Suppose they are unbiased with variances $var(f_1)$ and $var(f_2)$ and suppose they are uncorrelated.
- Then if you take a weighted average

$$f = wf_1 + (1-w) f_2$$

- The variance of the average is

$$\operatorname{var}(f) = w^2 \operatorname{var}\left(f_1\right) + (1-w)^2 \operatorname{var}\left(f_2\right)$$

Equal weights

- If w=1/2 then

$$\operatorname{var}(f) = \frac{\operatorname{var}\left(f_1\right) + \operatorname{var}\left(f_2\right)}{4}$$

Optimal Weights

$$\text{var}(f) = w^2\sigma_1^2 + (1-w)^2\sigma_2^2$$

- Minimizing with respect to w, the optimal weight

$$w = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}$$

$$= \frac{\sigma_1^{-2}}{\sigma_1^{-2} + \sigma_2^{-2}}$$

- The weight on forecast 1 is inversely proportional to its variance

Multiple Forecasts

- In general, if you have forecasts f_1 ,…, f_M a forecast combination is

$$f = w_1 f_1 + w_2 f_2 + \cdots + w_M f_M$$

- Where the weights are non-negative and

$$w_1 + w_2 + \cdots + w_M = 1$$

Optimal weights

- When the forecasts are uncorrelated, the optimal weights are

$$w_m = \frac{\sigma_m^{-2}}{\sigma_1^{-2} + \sigma_2^{-2} + \cdots + \sigma_M^{-2}}$$

- The weight on the m’th forecast is inversely proportional to its variance
- If they have the same variance, then the weights are all equal

Bates-Granger Combination

- Bates and Granger (1969)
  - An early influential paper
  - Suggested using empirical weights based on out-of-sample forecast variances

$$w_m = \frac{\hat{\sigma}_m^{-2}}{\hat{\sigma}_1^{-2} + \hat{\sigma}_2^{-2} + \cdots + \hat{\sigma}_M^{-2}}$$

  - Even though this was derived under the assumption of uncorrelated forecasts, this method can work well in practice.

Bates-Granger Implementation

- Take a series of (pseudo) out-of-sample forecasts and forecast errors
- Compute forecast variance (square of RMSE)
- Invert.
- Normalize by sum across all models

Example

RMSE	junk	yield	lags	
	0	1	2	3
AR(1)	2.46	2.38	2.34	2.34
AR(2)	2.46	2.37	2.32	2.32
AR(3)	2.41	2.33	2.36	2.37

- Take the first model with RMSE=2.46
- Square and invert to find 0.16
- Sum across all 12 models is 2.14
- Divide 0.16/2.14=0.08
- This is the weight for this model/forecast
- Because the RMSE is similar across models, the weights are very similar, all 0.08 or 0.09
- Bates-Granger weights essentially are the same as equal weights

Granger-Ramanathan Combination

- Granger and Ramanathan (1984)
- Introduced a regression method to combine forecasts
- Similar to a Mincer-Zarnowitz regression
- Regress the actual value on the forecasts
- Two forecasts:

$$y_t = \beta_1 f_{1t} + \beta_2 f_{2t} + e_t$$

Multiple Forecasts

$$y_t = \beta_1 f_{1t} + \beta_2 f_{2t} + \cdots + \beta_M f_{Mt} + e_t$$

- Should use a constrained regression
  - Omit intercept
  - Enforce non-negative coefficients
  - Constrain coefficients to sum to one

STATA implementation

- reg option **noconstant** removes the intercept
- Constrained regression command **cnsreg** enforces linear constraints defined by **constraint**
- For example, if you regress gdp on (p_1,p_2,p_3,p_4)
- **.constraint 1 p1+p2+p3+p4=1**
- **.cnsreg gdp p1 p2 p3 p4, constraints(1) noconstant**

Non-negativity

- In STATA it is difficult to enforce the non-negative condition on the weights
- You can do this manually
  - Estimate the regression
  - Eliminate a forecast with the most negative weight
  - Re-estimate
  - Keep eliminating forecasts until only positive weights are found.
- Another problem
  - If the forecasts are highly correlated, STATA may exclude redundant forecasts
  - That is okay, they were not helping anyway.

Example

```
. reg gdp p1 p2 p3 p4 p5 p6 p7 p8 p9 p10 p11 p12, noconstant
note: p2 omitted because of collinearity
note: p3 omitted because of collinearity
note: p4 omitted because of collinearity
note: p5 omitted because of collinearity
note: p8 omitted because of collinearity
note: p10 omitted because of collinearity

      Source |       SS       df       MS              Number of obs =      60
-------------+------------------------------           F(  6,    54) =   20.13
       Model |  580.076891     6  96.6794819           Prob > F      =  0.0000
    Residual |  259.363104    54  4.80302044           R-squared     =  0.6910
-------------+------------------------------           Adj R-squared =  0.6567
       Total |  839.439995    60  13.9906666           Root MSE      =  2.1916

------------------------------------------------------------------------------
         gdp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          p1 |  -2.011758   1.285095    -1.57   0.123    -4.588218     .564703
          p2 |  (omitted)
          p3 |  (omitted)
          p4 |  (omitted)
          p5 |  (omitted)
          p6 |   2.319188   1.252291     1.85   0.070    -.1915046    4.829881
          p7 |   .4674386   2.565805     0.18   0.856    -4.676691    5.611568
          p8 |  (omitted)
          p9 |  -.1637144    2.78425    -0.06   0.953      -5.7458    5.418371
         p10 |  (omitted)
         p11 |   1.232494   2.661173     0.46   0.645    -4.102837    6.567825
         p12 |  -1.032479   2.707502    -0.38   0.704    -6.460693    4.395735
------------------------------------------------------------------------------
```

Example

```
. constraint 1 p1+p6+p7+p9+p11+p12=1

. cnsreg gdp p1 p6 p7 p9 p11 p12, constraints(1) noconstant

Constrained linear regression                   Number of obs    =        60
                                                Root MSE         =    2.2429

 ( 1)  p1 + p6 + p7 + p9 + p11 + p12 = 1
```

gdp	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
p1	-.9284845	1.178581	-0.79	0.434	-3.290413	1.433444
p6	1.387255	1.17916	1.18	0.244	-.9758353	3.750346
p7	-1.858702	2.307476	-0.81	0.424	-6.482987	2.765583
p9	2.234642	2.539583	0.88	0.383	-2.854797	7.324081
p11	3.530808	2.42568	1.46	0.151	-1.330363	8.39198
p12	-3.365519	2.469361	-1.36	0.178	-8.314229	1.583192

```
. constraint 1 p6+p9=1

. cnsreg gdp p6 p9, constraints(1) noconstant

Constrained linear regression                   Number of obs    =        60
                                                Root MSE         =    2.2396

 ( 1)  p6 + p9 = 1
```

gdp	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
p6	.5196649	.1549455	3.35	0.001	.2096197	.82971
p9	.4803351	.1549455	3.10	0.003	.17029	.7903803

Granger-Ramanathan Weights and Forecast

- We found the following estimated weights
  - Model 6: 0.52
  - Model 9: 0.48
- Combination Forecast
  - 0.52*4.1+0.48*5.3=4.7%

Bayesian Model Averaging

- In our discussion of model selection, we pointed out that Bayes theorem says that when there are a set of models, one of which is true, then the probability that a model is true given the data is

$$P(M_1 \mid D) \propto \exp\left(-\frac{BIC}{2}\right)$$

- These can be used for forecast weights
- This is a simplified form of Bayesian model averaging (BMA) which is very popular

BMA formula

- We can write the weights as follows
- Let BIC* be the smallest BIC
  - The BIC of the best-fitting model
- Let ΔBIC=BIC-BIC* be the "BIC difference"

$$w_m^* = \exp\left(-\frac{\Delta BIC_m}{2} \right)$$

$$w_m = \frac{w_m^*}{\sum_{m=1}^{M} w_m^*}$$

Implementation

- Compute BIC for each model
- Find best-fitting BIC*
- Compute difference ΔBIC and exp(-ΔBIC/2)
- Sum up all values and re-normalize

BIC	junk	yield	lags	
	0	1	2	3
AR(1)	578	580	566*	571
AR(2)	581	584	569	574
AR(3)	585	587	573	578

-ΔBIC/2	junk	yield	lags	
	0	1	2	3
AR(1)	-6	-7	0	-2.5
AR(2)	-7.5	-9	-1.5	-4
AR(3)	-11.5	-10.5	-3.5	-6

weight	junk	yield	lags	
	0	1	2	3
AR(1)	0.00	0.00	0.75	0.06
AR(2)	0.00	0.00	0.15	0.02
AR(3)	0.00	0.00	0.02	0.00

- BMA puts the most weight on the model with the smallest BIC
- It puts very little weight on a model which has a BIC value quite different from the minimum
- In some cases, several models receive similar weight
- In this example, most weight (75%) goes on the model with the AR(1) plus 2 lags of the junk spread
- 15% also on AR(2) plus 2 lags

BMA Weights and Forecast

- BMA Forecast
  - 0.75*5.2+0.15*5.1+.02*5.3+.06*4.7+.02*4.3
    =5.1%

Weighted AIC (WAIC)

- Some authors have suggested replacing BIC with AIC in the weight formula

$$w_m \propto \exp\left(-\frac{AIC}{2}\right)$$

- There is not a strong theoretical foundation for this suggestion
- But, it is simple and works quite well in practice.

WAIC formula

- Let AIC* be the smallest AIC
  - The AIC of the best-fitting model
- ΔAIC=AIC-AIC* is the "AIC difference"

$$w_m^* = \exp\left(-\frac{\Delta AIC_m}{2}\right)$$

$$w_m = \frac{w_m^*}{\sum_{m=1}^{M} w_m^*}$$

AIC	junk	yield	lags	
	0	1	2	3
AR(1)	571	570	552*	554
AR(2)	571	571	552	554
AR(3)	571	570	552	554

-ΔAIC/2	junk	yield	lags	
	0	1	2	3
AR(1)	-8.5	-8	0	-1
AR(2)	-8.5	-8.5	0	-1
AR(3)	-8.5	-8	0	-1

weight	junk	yield	Lags	
	0	1	2	3
AR(1)	0.00	0.00	0.24	0.09
AR(2)	0.00	0.00	0.24	0.09
AR(3)	0.00	0.00	0.24	0.09

- WAIC splits weight more than BMA
- It puts 24% on each of the three models with the best near-equivalent AIC
- Puts positive weight on 6 models
- Puts zero weight on 6 models

WAIC Forecast

- WAIC Forecast
  - .24*5.2+.24*5.1+.24*5.3
    +.09*4.7+.09*4.3+.09*4.4
    =4.95%

Advantages of Combination Methods

- When the selection criterion (AIC, BIC) are very close for competing models, it is troubling to select one over the other based on a small different
  - In this setting WAIC and BMA will give the two models near-equal weight
- If the selection criterion are different, simple averaging gives all models the same weight, which seems naïve.
  - In this setting WAIC and BMA will give the models different weight
  - And will give zero weight if the different is sufficiently large
    - If the difference in the criterion is above 10.

GDP Combination Forecasts

- AIC Selection: 5.1%
- BIC Selection: 5.2%
- Simple Average: 4.4%
- Bates-Granger combination: 4.4%
- Granger-Ramanathan combination: 4.7%
- BMA: 5.1%
- WAIC: 4.95%

Example: Unemployment Rate Estimated on 1950-1995

	AIC	AIC weights	BIC	BIC weights
AR(4)	-1792	0	-1771	.16
AR(5)	-1799	.005	-1774*	.74
AR(6)	-1800	.01	-1770	.10
AR(7)	-1798	.005	-1764	0
AR(8)	-1797	0	-1758	0
AR(9)	-1795	0	-1752	0
AR(10)	-1793	0	-1746	0
AR(11)	-1800	.01	-1748	0
AR(12)	-1799	.005	-1743	0
AR(13)	-1808*	.57	-1748	0
AR(14)	-1806	.21	-1742	0
AR(15)	-1804	.08	-1735	0
AR(16)	-1803	.05	-1760	0
AR(17)	-1802	.03	-1724	0
AR(18)	-1800	.01	-1718	0
AR(19)	-1799	.005	-1712	0
AR(20)	-1798	.005	-1708	0

Out-of-Sample RMSE 1996-2010

Method	RMSE
AIC	.145
BIC	.145
BMA	.145
WAIC	.145
Best Model (AR(12))	.143

Which should you use?

- Current research suggests that combination methods achieve lower MSFE than selection
  - BMA achieves lower MSFE than BIC
  - WAIC achieves lower MSFE than AIC
- Naïve combination (simple averaging) works quite well
  - But the other methods can do better
- WAIC works well in practice
  - Bates-Granger also works well in many settings

Forecast Intervals

- How do you construct intervals for a combination forecast?
- Do not combine forecast intervals
- Given the weights, you can construct the sequence of sample forecasts and forecast errors
- Use these errors as you have before to construct the forecast interval
  - Compute the RMSE of the combination forecast error