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 2017Q1 based on data up through 2016
- GDP growth rate for the four quarters of 2016

2016Q1	2016Q2	2016Q3	2016Q4
0.8%	1.4%	3.5%	1.9%

Models

- In p.s. #9, you estimated models for GDP
  - AR(3) plus 3 lags of *dt3*
  - AR(3) plus 3 lags of *dt12*
  - AR(3) plus 3 lags of *spread1*
  - AR(3) plus 3 lags of *spread2*
  - AR(3) plus 3 lags of *corporate*
- The last model had the best AIC
  - Let's examine that model, but reconsider the number of lags

AIC for different lag structures

	corporate	yield	lags	
	0	1	2	3
AR(0)	1356	1346	1326	1328
AR(1)	1326	1325	1311.4	1313
AR(2)	1325	1325	1310.6*	1312
AR(3)	1327	1326	1312	1313

- The model with 2 AR lags and 2 lags of *corporate* has the lowest AIC
- But several models have similar values of AIC
  - (1,2)
  - (3,2)
  - (2,3)
  - (1,3)
  - (3,3)

Forecasts

	corp	yield	lags	
	0	1	2	3
AR(0)	3.1	3.4	3.8	3.8
AR(1)	2.7	2.9	3.3	3.2
AR(2)	2.8	2.9	3.4*	3.2
AR(3)	2.9	3.1	3.5	3.3

- The point forecasts are quite different
- The forecast selected by AIC is higher than the forecast from the AR model
- The models with 2 or lags of *corporate* have higher forecasts than the others

Average Forecast

- The average of the 16 forecasts is

$$\hat{y}_{average} = (3.1 + 2.7 + 2.8 + 2.9 + 3.4 + 2.9 + 2.9 + 3.1$$
$$+ 3.8 + 3.3 + 3.4 + 3.5 + 3.9 + 3.2 + 3.2 + 3.3)/16$$
$$= 3.2$$

- This is similar to a consensus or "Blue Chip" forecast.
- You could imagine these 16 forecasts as coming from different forecasters.
- Is it useful to combine the forecasts?

Pseudo Out-of-Sample Experiment

- Split the sample
  - Estimation period: 1954Q2-2004Q4 (50 years)
  - Evaluation period: 2005Q1-2014Q4 (10 years)
- Estimate the 16 models using 1954Q2-2004Q4
  - Fix the parameter estimates
- Use these models to forecast 2005Q1-2016Q4
- Also, take the average forecast for each period
- Create out-of-sample errors for the 16 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	corp	yield	lags	
	0	1	2	3
AR(0)	2.49	2.26	2.68	2.72
AR(1)	2.30	2.23*	2.55	2.54
AR(2)	2.29	2.24	2.52	2.50
AR(3)	2.30	2.24	2.65	2.62

RMSE	Average forecast
	2.18

- The comparisons based on out-of-sample RMSE are analogous to AIC on full sample, except the best models have only one lag of the corporate spread
- The lowest RMSE is **2.23**, achieved by the model with 1 lags of each
- But the RMSE of the average forecasts (the average across all 16 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

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

Equal weights

- If w=1/2 then

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

Optimal Weights

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

- 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	corp	yield	lags	
	0	1	2	3
AR(0)	2.49	2.26	2.68	2.72
AR(1)	2.30	2.23*	2.55	2.54
AR(2)	2.29	2.24	2.52	2.50
AR(3)	2.30	2.24	2.65	2.62

AR(0)	0.06	0.07	0.05	0.05
AR(1)	0.07	0.07	0.06	0.06
AR(2)	0.07	0.07	0.06	0.06
AR(3)	0.07	0.07	0.05	0.05

- Take the first model with RMSE=2.49
- Square and invert to find 0.16
- Sum across all 16 models is 2.70
- Divide 0.16/2.7=0.06
- This is the weight for this model/forecast
- Because the RMSE is similar across models, the weights are very similar, all in (0.05, 0.07)
- Bates-Granger weights close to 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 (out of sample) 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
- Apply to out-of-sample forecasts, not in-sample fitted values

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**
- Use out-of-sample dates
  - If the forecasts p1, p2, etc are only defined out of sample, this will is automatric

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.

```
. reg gdp y00 y10 y20 y30 y01 y11 y21 y31 y02 y12 y22 y32 y03 y13 y23 y33, noconstant
note: y10 omitted because of collinearity
note: y20 omitted because of collinearity
note: y30 omitted because of collinearity
note: y11 omitted because of collinearity
note: y21 omitted because of collinearity
note: y02 omitted because of collinearity
note: y12 omitted because of collinearity
note: y22 omitted because of collinearity
note: y32 omitted because of collinearity
```

Source	SS	df	MS
Model	214.680807	7	30.6686866
Residual	199.379193	41	4.86290716
Total	414.06	48	8.62625

Number of obs	=	48
F(7, 41)	=	6.31
Prob > F	=	0.0000
R-squared	=	0.5185
Adj R-squared	=	0.4363
Root MSE	=	2.2052

gdp	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
y00	-1.254081	.7183233	-1.75	0.088	-2.704764	.1966024
y10	0	(omitted)				
y20	0	(omitted)				
y30	0	(omitted)				
y01	1.675819	1.405378	1.19	0.240	-1.1624	4.514037
y11	0	(omitted)				
y21	0	(omitted)				
y31	-.2443925	.9002838	-0.27	0.787	-2.062553	1.573768
y02	0	(omitted)				
y12	0	(omitted)				
y22	0	(omitted)				
y32	0	(omitted)				
y03	.1560554	.6604141	0.24	0.814	-1.177678	1.489789
y13	.2061986	2.422056	0.09	0.933	-4.685242	5.09764
y23	1.199885	3.566298	0.34	0.738	-6.0024	8.40217
y33	-1.156802	2.15595	-0.54	0.594	-5.510831	3.197228

- 9 models omitted due to collinearity
- Re-estimate omitting these variables, imposing the constraint that the coefficients sum to 1

```
. constraint 1 y00+y01+y31+y03+y13+y23+y33=1

. cnsreg gdp y00 y01 y31 y03 y13 y23 y33, constraints(1) noconstant
note: y00 omitted because of collinearity

Constrained linear regression                      Number of obs     =         48
                                                   Root MSE          =     2.3447

 ( 1)  o.y00 + y01 + y31 + y03 + y13 + y23 + y33 = 1
```

gdp	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
y00	0	(omitted)				
y01	-.9611127	.5869497	-1.64	0.109	-2.14481	.2225842
y31	1.358167	.6685279	2.03	0.048	.009952	2.706382
y03	.5821505	.6526181	0.89	0.377	-.7339794	1.89828
y13	-3.083629	2.224263	-1.39	0.173	-7.569283	1.402025
y23	7.835886	2.787523	2.81	0.007	2.21431	13.45746
y33	-4.731462	1.82236	-2.60	0.013	-8.4066	-1.056323

- First model omitted due to collinearity
- Three models have negative weights
  - Omit last one, as it has most negative coefficient

```
. constraint 2 y01+y31+y03+y13+y23=1

. cnsreg gdp y01 y31 y03 y13 y23, constraints(2) noconstant

Constrained linear regression                   Number of obs     =         48
                                                Root MSE          =     2.4930

 ( 1)  y01 + y31 + y03 + y13 + y23 = 1
```

gdp	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
y01	-.3565666	.5728606	-0.62	0.537	-1.511091	.797958
y31	.9541807	.691284	1.38	0.174	-.4390106	2.347372
y03	-.2966184	.5932759	-0.50	0.620	-1.492287	.8990505
y13	-2.215773	2.338066	-0.95	0.348	-6.927834	2.496289
y23	2.914777	2.173206	1.34	0.187	-1.465032	7.294586

- Three models have negative weights
- Omit y13 as it is most negative

```
. constraint 3 y01+y31+y03+y23=1

. cnsreg gdp y01 y31 y03 y23, constraints(3) noconstant

Constrained linear regression                   Number of obs     =         48
                                                Root MSE          =     2.4901

 ( 1)  y01 + y31 + y03 + y23 = 1
```

gdp	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
y01	-.3464748	.5721129	-0.61	0.548	-1.498769	.8058198
y31	.9710667	.6902716	1.41	0.166	-.4192117	2.361345
y03	-.5763447	.5140567	-1.12	0.268	-1.611708	.4590186
y23	.9517528	.6567216	1.45	0.154	-.3709525	2.274458

- Two models have negative weights
- Omit y03 as it is most negative

```
. constraint 4 y01+y31+y23=1

. cnsreg gdp y01 y31 y23, constraints(4) noconstant

Constrained linear regression                    Number of obs     =         48
                                                 Root MSE          =     2.4971

 ( 1)  y01 + y31 + y23 = 1
```

gdp	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
y01	-.6575176	.5017432	-1.31	0.197	-1.667474	.352439
y31	1.407543	.5716051	2.46	0.018	.2569616	2.558124
y23	.2499746	.199275	1.25	0.216	-.1511452	.6510944

- One model has negative weights
- Omit it and re-estimate

Weights

```
. constraint 5 y31+y23=1

. cnsreg gdp y31 y23, constraints(5) noconstant

Constrained linear regression                   Number of obs     =         48
                                                Root MSE          =     2.5161

 ( 1)  y31 + y23 = 1
```

gdp	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
y31	.7039584	.1976411	3.56	0.001	.3063558	1.101561
y23	.2960416	.1976411	1.50	0.141	-.101561	.6936442

- Only two models receive positive weight
  - (3,1) gets 0.70
  - (2,3) gets 0.30

Granger-Ramanathan Weights and Forecast

- Combination Forecast
  - 0.7*3.1+0.3*3.2=3.1%
- Combination root MSE
  - 2.16
  - Lower than the 2.23 from single best model
  - Lower than the 2.18 from simple average

```
. gen e = gdp - y31*.7 - y23*.3
(231 missing values generated)

. summarize e

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
           e |        48   -1.277302    2.159734  -9.928538   1.992714
```

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	corp	yield	lags	
	0	1	2	3
AR(0)	1359	1353	1337	1342
AR(1)	1333	1335	1325*	1330
AR(2)	1336	1339	1328	1333
AR(3)	1341	1344	1333	1338

-ΔBIC/2	corp	yield	lags	
AR(0)	-17	-16	-6	-8.5
AR(1)	-4	-5	0	-2.5
AR(2)	-5.5	-7	-1.5	-4
AR(3)	-8	-9.5	-4	-6.5

weight	corp	yield	lags	
AR(0)	0.00	0.00	0.00	0.00
AR(1)	0.01	0.01	0.73	0.06
AR(2)	0.00	0.00	0.16	0.01
AR(3)	0.00	0.00	0.01	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 (73%) goes on the model with the AR(1) plus 2 lags of the corporate spread
- 16% also on AR(2) plus 2 lags
- 6% on AR(1) plus 3 lags

BMA Weights and Forecast

- BMA Forecast
  - 0.73*3.3+0.16*3.4+0.06*3.2+.01*(2.7+2.9+3.2+3.5)
    =3.3%

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”

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

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

	corp	yield	lags	
	0	1	2	3
AR(0)	1356	1346	1326	1328
AR(1)	1326	1325	1311.4	1313
AR(2)	1325	1325	1310.6*	1312
AR(3)	1327	1326	1312	1313

-ΔAIC/2	corp	yield	lags	
AR(0)	-22	-17	-7.5	-9
AR(1)	-7.5	-7	-0.4	-1
AR(2)	-7	-7	0	-0.5
AR(3)	-8	-7.5	-0.5	-1

weight	corp	yield	Lags	
AR(0)	0.00	0.00	0.00	0.00
AR(1)	0.00	0.00	0.19	0.10
AR(2)	0.00	0.00	0.28	0.17
AR(3)	0.00	0.00	0.17	0.10

- WAIC splits weight more than BMA
- It puts 17-28% on each of the four models with the best near-equivalent AIC
- Puts positive weight on 6 models
- Puts zero weight on 10 models

WAIC Forecast

- WAIC Forecast
  - 0.28*3.4+0.19*3.3+0.17*3.2
    +0.17*3.2+0.10*3.2+0.10*3.3
    =3.3%

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: 3.4%
- BIC Selection: 3.2%
- Simple Average: 3.2%
- Bates-Granger combination: 3.2%
- Granger-Ramanathan combination: 3.1%
- BMA: 3.3%
- WAIC: 3.3%

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

Assignments

- Diebold, Chapters 10, 11
- Read Chapter 11 from *The Signal and the Noise*
  - Last Reading Reflection
  - Thursday (4/20)
- Projects!