Forecasting
Lecture 2: Forecast Combination, Multi-Step Forecasts

Bruce E. Hansen

THE UNIVERSITY of WISCONSIN MADISON

Central Bank of Chile
October 29-31, 2013

Today's Schedule

- Combination Forecasts
- Multi-Step Forecasting
- Fan Charts
- Iterated Forecasts
- If time (optional): Threshold models or Nonlinear/Nonparametric Time Series

Review

- Optimal point forecast of y_{n+1} given information I_n is the conditional mean $E\left(y_{n+1}|I_n\right)$
- Linear model $E\left(y_{n+1}|I_n\right) \simeq \boldsymbol{\beta}'\mathbf{x}_n$ is an approximation
- Estimate linear projections by least-squares
- Model selection should focus on performance, not "truth"
  - Best forecast has smallest MSFE
  - Unknown, but MSFE can be estimated
  - CV is a good estimator of MSFE
- Good forecasts rely on selection of leading indicators

Combination Forecasts

Diversity of Forecasts

- Model choice is critical
  - Classic approach: Selection
  - Modern approach: Combination
- Issues:
  - How to select from a wide set of models/forecasts?
    - Model selection criteria
  - How to combine a wide set of models/forecasts?
    - Weight selection criteria

Foundation

- The ideal point forecast minimizes the MSFE
- The goal of a good combination forecast is to minimize the MSFE

Forecast Selection

- M forecasts: $\mathbf{f} = \{f(1), f(2), ..., f(M)\}$
- Selection picks $\hat{m}$ to determine the forecast $f = f(\hat{m})$
- M weights: $\mathbf{w} = \{w(1), w(2), ..., w(M)\}$
- A combination forecast is the weighted average

$$\begin{aligned} f(\mathbf{w}) &= \sum_{m=1}^{M} w(m) f(m) \\ &= \mathbf{w}'\mathbf{f} \end{aligned}$$

- Combination generalizes selection

Possible restrictions on the weight vector

- $\sum_{m=1}^{M} w(m) = 1$
  - Unbiasedness
  - Typically improves performance
- $w(m) \geq 0$
  - nonnegativity
  - regularization
  - Often critical for good performance
- $w(m) \in \{0, 1\}$
  - Equivalent to forecast selection
  - $f(\mathbf{w}) = f(m)$
  - Selection is a special case of combination
  - Strong restriction

OOS Forecast Combination

- Sequence of true out-of-sample forecasts $\mathbf{f}_t$ for y_{t+1}
- Combination forecast is $f(\mathbf{w}) = \mathbf{w}'\mathbf{f}$
- OOS empirical MSFE

$$\hat{\sigma}^2(\mathbf{w}) = \frac{1}{P} \sum_{t=n-P}^{n} \left(y_{t+1} - \mathbf{w}'\mathbf{f}_t \right)^2$$

- PLS selected the model with the smallest OOS MSFE
- Granger-Ramanathan combination: select $\mathbf{w}$ to minimize the OOS MSFE
- Minimization over $\mathbf{w}$ is equivalent to the least-squares regression of y_t on the forecasts

$$y_{t+1} = \mathbf{w}'\mathbf{f}_t + \varepsilon_{t+1}$$

Granger-Ramanathan (1984)

- Unrestricted least-squares

$$\hat{\mathbf{w}} = \left(\sum_{t=n-P}^{n} \mathbf{f}_t \mathbf{f}_t' \right)^{-1} \sum_{t=n-P}^{n} \mathbf{f}_t y_{t+1}$$

- This can produce weights far outside $[0, 1]$ and don't sum to one
- Granger-Ramanathan's intuition was that this flexibility is good
  - But they provided no theory to support conjecture
- Unrestricted weights are not regularized
  - This results in poor sampling performance

Alternative Representation

- Take $y_{t+1} = \mathbf{w}'\mathbf{f}_t + \varepsilon_{t+1}$, subtract y_{t+1} from each side

$$0 = \mathbf{w}'\mathbf{f}_t - y_{t+1} + \varepsilon_{t+1}$$

- Impose restriction that weights to sum to one.

$$0 = \mathbf{w}'\left(\mathbf{f}_t - y_{t+1}\right) + \varepsilon_{t+1}$$

- Define $\mathbf{e}_{t+1} = \mathbf{w}'\left(\mathbf{f}_t - y_{t+1}\right)$, the (negative) forecast errors. Then

$$0 = \mathbf{w}'\mathbf{e}_{t+1} + \varepsilon_{t+1}$$

- This is the regression of 0 on the forecast errors
- But it is still better to also impose non-negativity $w(m) \geq 0$

Constrained Granger-Ramanathan

The constrained GR weights solve the problem

$$\min_{\mathbf{w}} \mathbf{w}'\mathbf{A}\mathbf{w}$$

subject to

$$\sum_{m=1}^{M} w(m) = 1$$

$$0 \leq w(m) \leq 1$$

where

$$\mathbf{A} = \sum_{t} \mathbf{e}_{t+1}\mathbf{e}_{t+1}'$$

is the $M \times M$ matrix of forecast error empirical variances/covariances

Quadratic Programming (QP)

- The weights lie on the unit simplex
- The constrained GR weights minimize a quadratic over the unit simplex
- QP algorithms easily solve this problem
  - Gauss (qprog)
  - Matlab (quadprog)
  - R (quadprog)
- Solution solution typical
  - Many forecasts will receive zero weight

Bates-Granger (1969)

- Assume $\mathbf{A} = \sum_t \mathbf{e}_{t+1}\mathbf{e}'_{t+1}$ is diagonal.
- Then the regression with the coefficients constrained to sum to one

$$0 = \mathbf{w}'\mathbf{e}_{t+1} + \varepsilon_{t+1}$$

has solution

$$w(m) = \frac{\hat{\sigma}^{-2}(m)}{\sum_{j=1}^{M} \hat{\sigma}^{-2}(j)}$$

- This are the Bates-Granger weights.
- In many cases, they are close to equality, since OOS forecast variances can be quite similar

Bayesian Model Averaging (BMA)

- Put priors on individual models, and priors on the probability that model m is the true model
- Compute posterior probabilites $w(m)$ that m is the true model
- Forecast combination using $w(m)$
- Advantages
  - Conceptually simple
  - no theoretical analysis required
  - applies in broad contexts
- Disadvantages
  - Not designed to minimize forecast risk
  - Similar to BIC: asymptotically picks "true" finite models
  - does not distinguish between 1-step and multi-step forecast horizons

BMA Approximation

- BIC weights

$$w(m) \propto \exp\left(-\frac{BIC(m)}{2}\right)$$

- Simple approximation to full BMA method
- Smoothed version of BIC selection
- Works better than BIC selection in simulations

AIC Weights

- Smooted AIC

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

- Proposed by Buckland, Burnhamm and Augustin (1997)
- Not theoretically motivated, but works better than AIC selection in simulations

Comments

- Combination methods typically work better (lower MSFE) than comparable selection methods
- BIC and BMA not optimal for MSFE
- Granger-Ramanathan has similar senstive as PLS to choice of P
- Bates-Granger and weighted AIC have no theoretical grounding

Forecast Combination

$$
\begin{aligned}
\widehat{y}_{n+1}(\mathbf{w}) &= \sum_{m=1}^{M} w(m)\widehat{y}_{n+1}(m) \\
&= \sum_{m=1}^{M} w(m)\mathbf{x}_n(m)'\widehat{\beta}(m) \\
&= \mathbf{x}_n'\widehat{\beta}(\mathbf{w})
\end{aligned}
$$

where

$$
\widehat{\beta}(\mathbf{w}) = \sum_{m=1}^{M} w(m)\widehat{\beta}(m)
$$

- In linear models, the combination forecast is the same as the forecast based on the weighted average of the parameter estimates across the different models
- Computationally, it is easiest to calculate the M individual forecast $\hat{y}_{n+1}(m)$, then take the weighted average to obtain $\hat{y}_{n+1}(\mathbf{w})$

Combination Residuals

$$
\begin{aligned}
\widehat{e}_{t+1}(\mathbf{w}) &= y_{t+1} - \mathbf{x}_t' \widehat{\beta}(\mathbf{w}) \\
&= \sum_{m=1}^{M} w(m) \left(y_{t+1} - \mathbf{x}_t' \widehat{\beta}(m) \right) \\
&= \sum_{m=1}^{M} w(m) \widehat{e}_{t+1}(m)
\end{aligned}
$$

- In linear models, the residual from the combination model is the same as the weighted average of the model residuals.

Mallows Averaging Criterion

$$C_n(\mathbf{w}) = \widehat{\sigma}^2(\mathbf{w}) + \frac{2}{n}\widetilde{\sigma}^2 \sum_{m=1}^{M} w(m)k(m)$$

with $\widetilde{\sigma}^2$ an estimate from a "large" model

- $C_n(\mathbf{w})$ is an estimate of the MSFE (assuming homoskedasticity)
- Hansen (2007, Econometrica) Mallows Model Averaging (MMA)
- Hansen (Journal of Econometrics, 2008) Forecast Model Averaging (FMA)
- Combination weights found by constrained minimization

$$\widehat{\mathbf{w}} = \underset{\mathbf{w}}{\operatorname{argmin}}\, C_n(\mathbf{w})$$

subject to

$$\sum_{m=1}^{M} w(m) = 1$$

$$0 \leq w(m) \leq 1$$

- Solution by Quadratic Programming (QP)

Theory of Optimal Weights

- Hansen (2007, Econometrica)
- Mallows weight selection is asymptotically optimal under homoskedasticity
- [In large samples, equivalent to using MSFE-minimizing weights

Comparison of Granger-Ramanathan and FMA

- Both are solved by Quadratic Programming (QP)
- Both typically yield corner solutions – many forecasts will receive zero weight
- GR uses empirical (OOS) forecast errors, FMA uses sample residuals
- GR uses no penalty, FMA uses "average # of parameters" penalty
- FMA is an estimate of MSFE for homoskedastic one-step forecasts, GR has no optimality

Cross-Validation

- Leave-one-out estimator

$$\widehat{\beta}_{-t}(\mathbf{w}) = \sum_{m=1}^{M} w(m)\widehat{\beta}_{-t}(m)$$

- Leave-one-out prediction residual

$$\begin{aligned} \widetilde{e}_{t+1}(m) &= y_{t+1} - \sum_{m=1}^{M} w(m)\widehat{\beta}_{-t}(\mathbf{w})'\mathbf{x}_t(m) \\ &= \sum_{m=1}^{M} w(m)\widetilde{e}_{t+1}(m) \end{aligned}$$

- $CV_n(\mathbf{w}) = \frac{1}{n} \sum_{t=0}^{n-1} \widetilde{e}_{t+1}(\mathbf{w})^2$ is an estimate of $MSFE_n(m)$
- Cross-validation (CV) criterion for regression combination/averaging

Cross-validation Weights

Combination weights found by constrained minimization of $CV_n(\mathbf{w})$

$$\min_{\mathbf{w}} CV_n(\mathbf{w}) = \mathbf{w}'\widetilde{\mathbf{S}}\mathbf{w}$$

subject to

$$\sum_{m=1}^{M} w(m) = 1$$

$$0 \leq w(m) \leq 1$$

Cross-validation for combination forecasts (theory)

- **Theorem**: $ECV_n(\mathbf{w}) \simeq C_n(\mathbf{w})$
- For heteroskedastic forecasts, CV is a valid estimate of the one-step MSFE from a combination forecast
- Hansen and Racine (Journal of Econometrica, 2012) show that the CV weights are asymptotically optimal for cross-section data under heteroskedasticity

Summary: Forecast Combination Methods

- Granger-Ramanathan (GR), forecast model averaging (FMA) and cross-validation (CV) all pick weight vectors by quadratic minimization
- GR only needs actual forecasts, the method can be unknown or a black box
- CV can be computed for a wide variety of estimation methods
  - optimality theory for linear estimation
- FMA limited to homoskedastic one-step-ahead models
- Smoothed AIC (SAIC) and BMA have no forecast optimality, and are designed for homoskedastic one-step-ahead forecasts.

Example: AR models for GDP Growth

- Fit AR(1) and AR(2) only
- Leave-one-out residuals $\tilde{e}_{1t}$ and $\tilde{e}_{2t}$
- Covariance matrix
$$\widetilde{\mathbf{S}} = \begin{bmatrix} 10.72 & 10.44 \\ 10.44 & 10.52 \end{bmatrix}$$
- The best-fitting single model is AR(2)
- The best combination is $\mathbf{w} = (.22, .78)'$
- $CV = 10.50$

Example: AR models for GDP Growth

- Fit AR(0) through AR(12)
- AR(0) is constant only
- Models with positive weight are AR(0), AR(1), AR(2)
- $\mathbf{w} = (.06, .16, .78)'$

$$\widetilde{\mathbf{S}} = \begin{bmatrix} 12.0 & 10.6 & 10.4 \\ 10.6 & 10.7 & 10.4 \\ 10.4 & 10.5 & 10.5 \end{bmatrix}$$

- $CV = 10.50$ (essentially unchanged)

Example: Leading Indicator Forecasts

- Fit AR(1), AR(2) with leading indicators
- Models with positive weight

	w
AR(1), Spread, Housing	0.13
AR(1), Spread, High-Yield, Housing	0.16
AR(1), Spread, High-Yield, Housing, Building	0.52
AR(2)	0.18
AR(2), Spread	0.01

- $CV = 9.81$

Summary: Forecast Combination by CV

- M forecasts $\widehat{f}_{n+1}(m)$ from n observations
- For each estimate m
  - Define the leave-one-out prediction error

$$\begin{aligned}\widetilde{e}_{t+1}(m) &= y_{t+1} - \widehat{\beta}'_{(-t)}(m)\mathbf{x}_t(m) \\ &= \frac{\widehat{e}_{t+1}(m)}{1-h_{tt}(m)}\end{aligned}$$

  - Store the $n \times 1$ vector $\widetilde{\mathbf{e}}(m)$
- Construct the $M \times M$ matrix

$$\widetilde{\mathbf{S}} = \frac{1}{n}\widetilde{e}'\widetilde{e}$$

- Find the $M \times 1$ weight vector $\mathbf{w}$ which minimizes $\mathbf{w}'\widetilde{\mathbf{S}}\mathbf{w}$
  - Use quadratic programming (quadprog) to find solution
- The combination forecast is $\widehat{f}_{n+1} = \sum_{m=1}^{M} w(m)\widehat{f}_{n+1}(m)$

Forecast Combination Criticisms

- There has been considerable skepticism about formal forecast combination method in the forecast literature
- Many researchers have found that equal weighting: $(w_m = 1/M)$ works as well as formal methods
- However, the formal methods which investigated are
  - Bates-Granger simple weights
    - Not expected by theory to work well
  - Unconstrained Granger-Ramanathan
    - Without imposing $[0, 1]$ weights, work terribly!
- Furthermore, most investigations examine pseudo out-of-sample performance
  - Identical to comparing models by PLS criterion
  - This is NOT an investigation of performance
  - Just a ranking by PLS

Another Example - 10-Year Bond Rate

- Estimated AR(1) through AR(24) models
- CV Selection picked AR(2)
- CV weight Selection: Models with positive weight
  - AR(0): $w = 0.04$
  - AR(1): $w = 0.04$
  - AR(2): $w = 0.47$
  - AR(6): $w = 0.23$
  - AR(22): $w = 0.22$
- MInimizing $CV = 0.0761$ (slightly lower than 0.0768 from AR(2))
- Point forecast 1.96 (same as from AR(2))

Multi-Step Forecasts

Multi-Step Forecasts

- Forecast horizon: h
- We say the forecast is "multi-step" if $h > 1$
- Forecasting y_{n+h} given I_n
- e.g., forecasting GDP growth for 2012:3, 2012:4, 2013:1, 2013:2
- The forecast distribution is $y_{n+h} \mid I_n \sim F_h(y_{n+h}|I_n)$

Point Forecast

- $f_{n+h|h}$ minimizes expected squared loss

$$
\begin{aligned}
f_{n+h|h} &= \underset{f}{\operatorname{argmin}} \, E\left(\left(y_{n+h} - f \right)^2 | I_n \right) \\
&= E\left(y_{n+h} | I_n \right)
\end{aligned}
$$

- Optimal point forecasts are h-step conditional means

Relationship Between Forecast Horizons

- Take an AR(1) model

$$y_{t+1} = \alpha y_t + u_{t+1}$$

- Iterate

$$\begin{aligned} y_{t+1} &= \alpha \left(\alpha y_{t-1} + u_t \right) + u_{t+1} \\ &= \alpha^2 y_{t-1} + \alpha u_t + u_{t+1} \end{aligned}$$

or

$$\begin{aligned} y_{t+2} &= \alpha^2 y_t + e_{t+2} \\ u_{t+2} &= \alpha u_{t+1} + u_{t+2} \end{aligned}$$

- Repeat h times

$$\begin{aligned} y_{t+h} &= \alpha^h y_t + e_{t+h} \\ e_{t+h} &= u_{t+h} + \alpha u_{t+h-1} + \alpha^2 u_{t+h-2} + \cdots + \alpha^{h-1} u_{t+1} \end{aligned}$$

AR(1)

h-step forecast

$$\begin{aligned} y_{t+h} &= \alpha^h y_t + e_{t+h} \\ e_{t+h} &= u_{t+h} + \alpha u_{t+h-1} + \alpha^2 u_{t+h-2} + \cdots + \alpha^{h-1} u_{t+1} \\ E\left(y_{n+h} | I_n\right) &= \alpha^h y_n \end{aligned}$$

- $h-$step point forecast is linear in y_n
- h-step forecast error e_{n+h} is a MA$(h-1)$

AR(2) Model

- 1-step AR(2) model

$$y_{t+1} = \alpha_0 + \alpha_1 y_t + \alpha_2 y_{t-1} + u_{t+1}$$

- 2-steps ahead

$$y_{t+2} = \alpha_0 + \alpha_1 y_{t+1} + \alpha_2 y_t + u_{t+2}$$

- Taking conditional expectations

$$\begin{aligned} E\left(y_{t+2} \middle| I_t\right) &= \alpha_0 + \alpha_1 E\left(y_{t+1} \middle| I_t\right) + \alpha_2 E\left(y_t \middle| I_t\right) + E\left(e_{t+2} \middle| I_t\right) \\ &= \alpha_0 + \alpha_1\left(\alpha_0 + \alpha_1 y_t + \alpha_2 y_{t-1}\right) + \alpha_2 y_t \\ &= \alpha_0 + \alpha_1\alpha_0 + \left(\alpha_1^2 + \alpha_2\right) y_t + \alpha_1\alpha_2 y_{t-1} \end{aligned}$$

  which is linear in (y_t, y_{t-1})

- In general, a 1-step linear model implies an *h*-step approximate linear model in the same variables

AR(k) h-step forecasts

If

$$y_{t+1} = \alpha_0 + \alpha_1 y_t + \alpha_2 y_{t-1} + \cdots + \alpha_k y_{t-k+1} + u_{t+1}$$

then

$$y_{t+h} = \beta_0 + \beta_1 y_t + \beta_2 y_{t-1} + \cdots + \beta_k y_{t-k+1} + e_{t+h}$$

where e_{t+h} is a MA(h-1)

Leading Indicator Models

If

$$y_{t+1} = \mathbf{x}_t' \beta + u_t$$

then

$$E\left(y_{t+h} | I_t\right) = E\left(\mathbf{x}_{t+h-1} | I_t\right)' \beta$$

If $E\left(\mathbf{x}_{t+h-1} | I_t\right)$ is itself (approximately) a linear function of $\mathbf{x}_t$, then

$$E\left(y_{t+h} | I_t\right) = \mathbf{x}_t' \gamma$$

$$y_{t+h} = \mathbf{x}_t' \gamma + e_{t+h}$$

Common Structure: h-step conditional mean is similar to 1-step structure, but error is a MA.

Forecast Variable

- We should think carefully about the variable we want to report in our forecast
- The choice will depend on the context
- What do we want to forecast?
  - Future level: y_{n+h}
    - interest rates, unemployment rates
  - Future differences: Δy_{t+h}
  - Cummulative Change: Δy_{t+h}
    - Cummulative GDP growth

Forecast Transformation

- $f_{n+h|n} = E\left(y_{n+h}|I_n\right) =$ expected future level
  - Level specification

$$
\begin{aligned}
y_{t+h} &= \mathbf{x}_t'\boldsymbol{\beta} + e_{t+h} \\
f_{n+h|n} &= \mathbf{x}_t'\boldsymbol{\beta}
\end{aligned}
$$

  - Difference specification

$$
\begin{aligned}
\Delta y_{t+h} &= \mathbf{x}_t'\boldsymbol{\beta}_h + e_{t+h} \\
f_{n+h|n} &= y_n + \mathbf{x}_t'\boldsymbol{\beta}_1 + \cdots + \mathbf{x}_t'\boldsymbol{\beta}_h
\end{aligned}
$$

  - Multi-Step difference specification

$$
\begin{aligned}
y_{t+h} - y_t &= \mathbf{x}_t'\boldsymbol{\beta} + e_{t+h} \\
f_{n+h|n} &= y_n + \mathbf{x}_t'\boldsymbol{\beta}
\end{aligned}
$$

Direct and Iterated

- There are two methods of multistep $(h > 1)$ forecasts
- Direct Forecast
  - Model and estimate $E\left(y_{n+h}|I_n\right)$ directly
- Iterated Forecast
  - Model and estimate one-step $E\left(y_{n+1}|I_n\right)$
  - Iterate forward h steps
  - Requires full model for all variables
- Both have advantages and disadvantages
  - For now, we will forcus on direct method.

Direct Multi-Step Forecasting

- Markov approximation
  - $E\left(y_{n+h}|I_n\right) = E\left(y_{n+h}|x_n, x_{n-1}, ...\right) \approx E\left(y_{n+h}|x_n, ..., x_{n-p}\right)$
- Linear approximation
  - $E\left(y_{n+h}|x_n, ..., x_{n-p}\right) \approx \boldsymbol{\beta}'\mathbf{x}_n$
- Projection Definition
  - $\boldsymbol{\beta} = \left(E\left(\mathbf{x}_t\mathbf{x}_t'\right)\right)^{-1}\left(E\left(\mathbf{x}_t y_{t+h}\right)\right)$
- Forecast error
  - $e_{t+h} = y_{t+h} - \boldsymbol{\beta}'\mathbf{x}_t$

Multi-Step Forecast Model

$$y_{t+h} = \beta' \mathbf{x}_t + e_{t+h}$$

$$\begin{aligned} \beta &= \left(E\left(\mathbf{x}_t \mathbf{x}_t'\right)\right)^{-1} \left(E\left(\mathbf{x}_t y_{t+h}\right)\right) \\ E\left(\mathbf{x}_t e_{t+h}\right) &= 0 \\ \sigma^2 &= E\left(e_{t+h}^2\right) \end{aligned}$$

Least Squares Estimation

$$\widehat{\beta} = \left(\sum_{t=0}^{n-1} \mathbf{x}_t \mathbf{x}_t'\right)^{-1} \left(\sum_{t=0}^{n-1} \mathbf{x}_t y_{t+h}\right)$$

$$\widehat{y}_{n+h|n} = \widehat{f}_{n+h|n} = \widehat{\beta}' \mathbf{x}_n$$

Residuals

- Least-squares residuals
  - $\widehat{e}_{t+h} = y_{t+h} - \widehat{\beta}' \mathbf{x}_t$
  - Standard, but overfit
- Leave-one-out residuals
  - $\widetilde{e}_{t+h} = y_{t+h} - \widehat{\beta}'_{-t} \mathbf{x}_t$
  - Does not correct for MA errors
- Leave h out residuals

$$\widetilde{e}_{t+h} = y_{t+h} - \widehat{\beta}'_{-t,h} \mathbf{x}_t$$

$$\widehat{\beta}_{-t,h} = \left(\sum_{|j+h-t| \geq h} \mathbf{x}_j \mathbf{x}_j' \right)^{-1} \left(\sum_{|j+h-t| \geq h} \mathbf{x}_j y_{j+h} \right)$$

- The summation is over all observations outside $h-1$ periods of $t+h$.

Example: GDP Forecast

$$y_t = 400 \log(GDP_t)$$

Forecast Variable: GDP growth over next h quarters, at annual rate

$$\frac{y_{t+h} - y_t}{h} = \beta_0 + \beta_1 \Delta y_t + \beta_1 \Delta y_{t-1} + Spread_t + HighYield_t + \beta_2 HS_t + e_{t+h}$$

HS_t =Housing Starts$_t$

	$h=1$	$h=2$	$h=3$	$h=4$
β_0	-0.33 (1.0)	-0.38 (1.3)	-0.01 (1.6)	0.47 (1.8)
Δy_t	0.16 (.10)	0.18 (.09)	0.13 (.08)	0.13 (.09)
Δy_{t-1}	0.09 (.10)	0.04 (.05)	0.05 (.07)	0.02 (.06)
$Spread_t$	0.61 (.23)	0.65(.19)	0.65 (.22)	0.65 (.25)
$HighYield_t$	-1.10 (.75)	-0.68 (.70)	-0.48 (.90)	-0.41 (1.01)
HS_t	1.86 (.65)	1.64 (.70)	1.31 (.80)	1.01 (.94)

Example: GDP Forecast

	Cummulative Annualized Growth	
	Forecast	Actual
2012:2	1.3	1.2
2012:3	1.6	2.0
2012:4	2.9	1.4
2013:1	2.2	1.3
2013:2	2.4	1.5
2013:3	2.7	
2013:4	2.9	
2014:1	3.2	

Selection and Combination for h step forecasts

- AIC routinely used for model selection
- PLS (OOS MSFE) routinely used for model evaluation
- Neither well justified
- Not well studied problem
- I recommend "leave h out" cross-validation.
- Topic of afternoon seminar
- Minimize sum of squared "leave h out" residuals, separately for each forecast horizon.

Example: GDP Forecast Weights by Horizon

	$h=1$	$h=2$	$h=3$	$h=4$	$h=5$	$h=6$	$h=7$
AR(1)		.15	.19	.28	.18	.16	.11
AR(2)	.30						
AR(1)+HS	.66	.70	.22				
AR(1)+HS+BP		.14	.58	.72	.82	.84	.89
AR(2)+HS	.04						
$\widehat{y}_{n+h\|n}$	1.7	2.0	1.9	2.0	2.1	2.3	2.6

h-step Variance Forecasting

- Not well developed using direct methods

h-step Interval Forecasts

- Similar to 1-step interval forecasts
  - But calculated from $h-$step residuals
- Use constant variance specification
- Let $\widehat{q}^e(\alpha)$ and $\widehat{q}^e(1-\alpha)$ be the α'th and $(1-\alpha)$'th percentiles of residuals $\widetilde{e}_{t+h}$
- Forecast Interval:

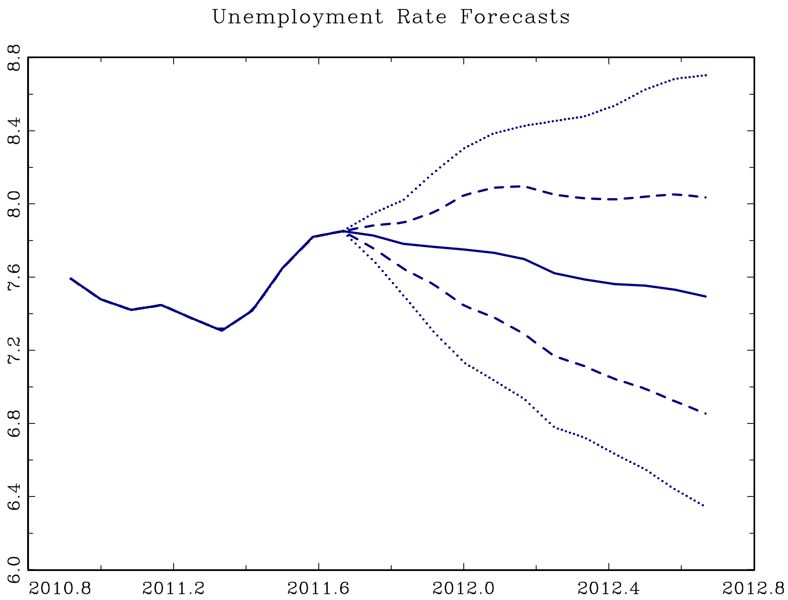
$$[\widehat{\mu}_n + \widehat{q}^\varepsilon(\alpha), \quad \widehat{\mu}_n + \widehat{q}^e(1-\alpha)]$$

Fan Charts

- Plots of a set of interval forecasts for multiple horizons
  - Pick a set of horizons, $h = 1, ..., H$
  - Pick a set of quantiles, e.g. $\alpha = .10, .25, .75, .90$
  - Recall the quantiles of the conditional distribution are $q_n(\alpha, h) = \mu_n(h) + \sigma_n(h) q^{\varepsilon}(\alpha, h)$
  - Plot $q_n(.1, h), q_n(.25, h), \mu_n(h), q_n(.75, h), q_n(.9, h)$ against h
- Graphs easier to interpret than tables

Illustration

- I've been making monthly forecasts of the Wisconsin unemployment rate
- Forecast horizon $h = 1, ..., 12$ (one year)
- Quantiles: $\alpha = .1, .25, .75, .90$
- This corresponds to plotting 50% and 80% forecast intervals
- 50% intervals show "likely" region (equal odds)

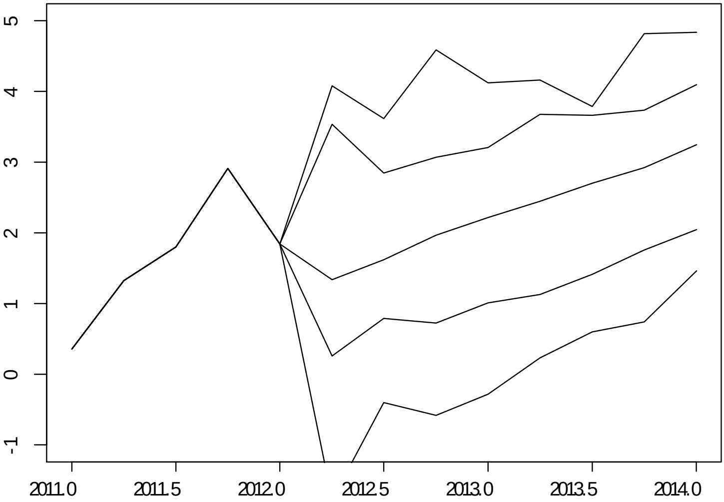

Unemployment Rate Forecasts
8.8
8.4
8.0
7.6
7.2
6.8
6.4
6.0
2010.8
2011.2
2011.6
2012.0
2012.4
2012.8

Comments

- Showing the recent history gives perspective
- Some published fan charts use colors to indicate regions, but do not label the colors
- Labels important to infer probabilities
- I like clean plots, not cluttered

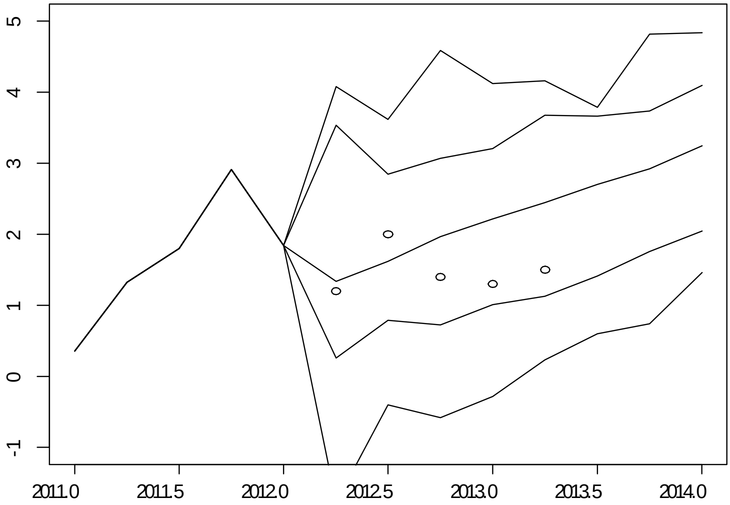
Illustration: GDP Growth

Figure: GDP Average Growth Fan Chart



It doesn't "fan" because we are plotting average growth

Figure: Fan Chart with Actuals



Iterated Forecasts

- Estimate one-step forecast
- Iterate to obtain multi-step forecasts
- Only works in complete systems
  - Autoregressions
  - Vector autoregressions

Vector Autoregresive Models

- $\mathbf{y}_t$ is an p vector
- x_t are other variables (including lags)
- Ideal point forecast $E\left(\mathbf{y}_{n+1}|I_n\right)$
- Linear approximation

$$E\left(\mathbf{y}_{n+1}|I_n\right) \simeq A_1\mathbf{y}_t + A_2\mathbf{y}_{t-1} + \cdots + A_k\mathbf{y}_{t-k+1} + Bx_t$$

- Vector Autoregression (VAR)

$$\mathbf{y}_{t+1} = A_1\mathbf{y}_t + A_2\mathbf{y}_{t-1} + \cdots + A_k\mathbf{y}_{t-k+1} + Bx_t + e_{t+1}$$

- Estimation: Least squares

$$\mathbf{y}_{t+1} = \widehat{A}_1\mathbf{y}_t + \widehat{A}_2\mathbf{y}_{t-1} + \cdots \widehat{+A_k}\mathbf{y}_{t-k+1} + \widehat{B}x_t + e_{t+1}$$

- One-Step-Ahead Point forecast

$$\widehat{\mathbf{y}}_{n+1} = \widehat{A}_1\mathbf{y}_n + \widehat{A}_2\mathbf{y}_{n-1} + \cdots + \widehat{A}_k\mathbf{y}_{n-k+1} + \widehat{B}x_n$$

Vector Autoregresive versus Univariate Models

- Let $\mathbf{x}_t = (\mathbf{y}_t, \mathbf{y}_{t-1}, ..., x_t)$
- Then a VAR is a set of p regression models

$$\begin{aligned} y_{1t+1} &= \beta_1' \mathbf{x}_t + e_{1t} \\ &\vdots \\ y_{pt+1} &= \beta_p' \mathbf{x}_t + e_{pt} \end{aligned}$$

- All variables $\mathbf{x}_t$ enter symmetrically in each equation
- Sims (1980) argued that there is no a priori reason to include or exclude an individual variable from an individual equation.

Model Selection

- Do not view selection as identification of "truth"
- Rather, inclusion/exclusion is to improve finite sample performance
  - minimize MSFE
- Use selection methods, equation-by-equation

Example: VAR with 2 variables

$$
\begin{aligned}
y_{1t+1} &= \widehat{\beta}_{11} y_{1t} + \widehat{\beta}_{12} y_{1t-1} + \widehat{\beta}_{13} y_{2t} + \widehat{e}_{1t} \\
&\vdots \\
y_{2t+1} &= \widehat{\beta}_{21} y_{1t} + \widehat{\beta}_{22} y_{2t} + \widehat{\beta}_{23} y_{2t-1} + \widehat{e}_{2t}
\end{aligned}
$$

- Selection picks y_{1t}, y_{1t-1}, y_{2t} for equation for y_{1t+1}
- Selection picks y_{1t}, y_{2t}, y_{2t-1} for equation for y_{2t+1}
- The two equations have different variables

- Same as system

$$\mathbf{y}_{t+1} = A_1 \mathbf{y}_t + A_2 \mathbf{y}_{t-1} + e_{t+1}$$

with

$$A_1 = \begin{bmatrix} \beta_{11} & \beta_{13} \\ \beta_{21} & \beta_{22} \end{bmatrix}$$

$$A_2 = \begin{bmatrix} \beta_{12} & 0 \\ 0 & \beta_{23} \end{bmatrix}$$

- The VAR system notation is still quite useful for many purposes (including multi-step forecasting)

Iterative Forecast Relationships in Linear VAR

- vector y_t

$$y_{t+1} = A_0 + A_1 y_t + A_2 y_{t-1} + \cdots + A_k y_{t-k+1} + u_{t+1}$$

- 1-step conditional mean

$$\begin{aligned} E\left(y_{t+1} | I_t\right) &= A_0 + A_1 E\left(y_t | I_t\right) + \cdots + A_k E\left(y_{t-k+1} | I_t\right) \\ &= A_0 + A_1 y_t + A_2 y_{t-1} + \cdots + A_k y_{t-k+1} \end{aligned}$$

- 2-step conditional mean

$$\begin{aligned} E\left(y_{t+1} | I_{t-1}\right) &= E\left(E\left(y_{t+1} | I_t\right) | I_{t-1}\right) \\ &= A_0 + A_1 E\left(y_t | I_{t-1}\right) + \cdots + A_k E\left(y_{t-k+1} | I_{t-1}\right) \\ &= A_0 + A_1 E\left(y_t | I_{t-1}\right) + A_2 y_{t-1} + \cdots + A_k y_{t-k+1} \end{aligned}$$

- $h-$step conditional mean

$$\begin{aligned} E\left(y_{t+1} | I_{t-h+1}\right) &= E\left(E\left(y_{t+1} | I_t\right) |_{I_{t-h+1}}\right) \\ &= A_0 + A_1 E\left(y_t | I_{t-h+1}\right) + \cdots + A_k E\left(y_{t-k+1} | I_{t-h+1}\right. \end{aligned}$$

- Linear in lower-order (up to $h-1$ step) conditional means

Iterative Least Squares Forecasts

- Estimate 1-step VAR(k) by least-squares

$$y_{t+1} = \widehat{A}_0 + \widehat{A}_1 y_t + \widehat{A}_2 y_{t-1} + \cdots + \widehat{A}_k y_{t-k+1} + \widehat{u}_{t+1}$$

- Gives 1-step point forecast

$$\widehat{y}_{n+1|n} = \widehat{A}_0 + \widehat{A}_1 y_n + \widehat{A}_2 y_{n-1} + \cdots + \widehat{A}_k y_{n-k+1}$$

- 2-step iterative forecast

$$\widehat{y}_{n+2|n} = \widehat{A}_0 + \widehat{A}_1 \widehat{y}_{n+1|n} + \widehat{A}_2 y_n + \cdots + \widehat{A}_k y_{n-k+2}$$

- $h-$step iterative forecast

$$\widehat{y}_{n+h|n} = \widehat{A}_0 + \widehat{A}_1 \widehat{y}_{n+h-1|n} + \widehat{A}_2 \widehat{y}_{n+h-2|n} + \cdots + \widehat{A}_k \widehat{y}_{n+h-k|n}$$

- This is (numerically) different than the direct LS forecast

Illustration 1: GDP Growth

- AR(2) Model
- $y_{t+1} = 1.6 + 0.30y_t + .16y_{t-1}$
- $y_n = 1.8,\ y_{n-1} = 2.9$
- $\widehat{y}_{n+1} = 1.6 + 0.30 * 1.8 + .16 * 2.9 = 2.6$
- $\widehat{y}_{n+2} = 1.6 + 0.30 * 2.6 + .16 * 1.8 = 2.7$
- $\widehat{y}_{n+3} = 1.6 + 0.30 * 2.7 + .16 * 2.6 = 2.9$
- $\widehat{y}_{n+4} = 1.6 + 0.30 * 2.9 + .16 * 2.7 = 3.0$

Point Forecasts

2012:2	2.65
2012:3	2.72
2012:4	2.87
2013:1	2.93
2013:2	2.97
2013:3	2.99
2013:4	3.00
2014:1	3.01

Illustration 2: GDP Growth+Housing Starts

- VAR(2) Model
- y_{1t} = GDP Growth, y_{2t} =Housing Starts
- x_t = (GDP Growth$_t$, Housing Starts$_t$, GDP Growth$_{t-1}$, Housing Starts$_{t-1}$
- $y_{t+1} = \widehat{A}_0 + \widehat{A}_1 y_t + \widehat{A}_2 y_{t-1} + \widehat{u}_{t+1}$
- $y_{1t+1} = 0.43 + 0.15 y_{1t} + 11.2 y_{2t} + 0.18 y_{1t-1} - 10.1 y_{2t-1}$
- $y_{2t+1} = 0.07 - 0.001 y_{1t} + 1.2 y_{2t} - 0.001 y_{1t-1} - 0.26 y_{2t-1}$

Illustration 2: GDP Growth+Housing Starts

- $y_{1n} = 1.8$, $y_{2n} = 0.71$, $y_{1n-1} = 2.9$, $y_{2n-1} = 0.68$
- $y_{1n+1} = 0.43 + 0.15 * 1.8 + 11.2 * 0.71 + 0.18 * 2.9 - 10.1 * 0.68 = 2.3$
- $y_{2t+1} = 0.07 - 0.001 * 1.8 + 1.2 * 0.71 - 0.001 * 2.9 - 0.26 * 0.68 = 0.76$
- $y_{1n+2} = 0.43 + 0.15 * 2.3 + 11.2 * 0.76 + 0.18 * 1.8 - 10.1 * 0.71 = 2.4$
- $y_{2t+1} = 0.07 - 0.001 * 2.3 + 1.2 * 0.76 - 0.001 * 1.8 - 0.26 * 0.71 = 0.80$

Point Forecasts

	GDP	Housing
2012:2	2.36	0.76
2012:3	2.38	0.80
2012:4	2.53	0.84
2013:1	2.58	0.88
2013:2	2.64	0.92
2013:3	2.66	0.95
2013:4	2.69	0.98
2014:1	2.71	1.01

Model Selection

- It is typical to select the 1-step model and use this to make all h-step forecasts
- However, there theory to support this is incomplete
- (It is not obvious that the best 1-step estimate produces the best h-step estimate)
- For now, I recommend selecting based on the 1-step estimates

Model Combination

- There is no theory about how to apply model combination to h-step iterated forecasts
- Can select model weights based on 1-step, and use these for all forecast horizons

Variance, Distribution, Interval Forecast

- While point forecasts can be simply iterated, the other features cannot
- Multi-step forecast distributions are convolutions of the 1-step forecast distribution.
  - Explicit calculation computationally costly beyond 2 steps
- Instead, simple simulation methods work well
- The method is to use the estimated condition distribution to simulate each step, and iterate forward. Then repeat the simulation many times.

Multi-Step Forecast Simulation

- Let $\mu(\mathbf{x})$ and $\sigma(\mathbf{x})$ denote the models for the conditional one-step mean and standard deviation as a function of the conditional variables $\mathbf{x}$
- Let $\widehat{\mu}(\mathbf{x})$ and $\widehat{\sigma}(\mathbf{x})$ denote the estimates of these functions, and let $\{\widehat{\varepsilon}_1, ..., \widehat{\varepsilon}_n\}$ be the normalized residuals
- $\mathbf{x}_n = (y_n, y_{n-1}, ..., y_{n-p})$ is known. Set $\mathbf{x}_n^* = \mathbf{x}_n$
- To create one h-step realization:
  - Draw ε_{n+1}^* iid from normalized residuals $\{\widehat{\varepsilon}_1, ..., \widehat{\varepsilon}_n\}$
  - Set $y_{n+1}^* = \widehat{\mu}(\mathbf{x}_n^*) + \widehat{\sigma}(\mathbf{x}_n^*)\,\varepsilon_{t+1}^*$
  - Set $\mathbf{x}_{n+1}^* = (y_{n+1}^*, y_n, ..., y_{n-p+1})$
  - Draw ε_{n+2}^* iid from normalized residuals $\{\widehat{\varepsilon}_1, ..., \widehat{\varepsilon}_n\}$
  - Set $y_{n+2}^* = \widehat{\mu}(\mathbf{x}_{n+1}^*) + \widehat{\sigma}(\mathbf{x}_{n+1}^*)\,\varepsilon_{t+2}^*$
  - Set $\mathbf{x}_{n+2}^* = (y_{n+2}^*, y_{n+1}^*, ..., y_{n-p+2})$
  - Repeat until you obtain y_{n+h}^*
  - y_{n+h}^* is a draw from the h step ahead distribution
- Repeat this B times, and let $y_{n+h}^*(b)$, $b = 1, ..., B$ denote the B repetitions

Multi-Step Forecast Simulation

- The simulation has produced $y^*_{n+h}(b)$, $b = 1, ..., B$
- For forecast intervals, calculate the empirical quantiles of $y^*_{n+h}(b)$
  - For an 80% interval, calculate the 10% and 90%
- For a fan chart
  - Calculate a set of empirical quantiles (10%, 25%, 75%, 90%)
  - For each horizon $h = 1, ..., H$
- As the calculations are linear they are numerically quick
  - Set B large
  - For a quick application, $B = 1000$
  - For a paper, $B = 10,000$ (minimum))

VARs and Variance Simulation

- The simulation method requires a method to simulate the conditional variances
- In a VAR setting, you can:
  - Treat the errors as iid (homoskedastic)
    - Easiest
  - Treat the errors as independent GARCH errors
    - Also easy
  - Treat the errors as multivariate GARCH
    - Allows volatility to transmit across variables
    - Probably not necessary with aggregate data