Time Series and Forecasting
Lecture 3
Forecast Intervals, Multi-Step Forecasting

Bruce E. Hansen

THE UNIVERSITY of WISCONSIN MADISON

Summer School in Economics and Econometrics
University of Crete
July 23-27, 2012

Today's Schedule

- Review
- Forecast Intervals
- Forecast Distributions
- Multi-Step Direct Forecasts
- Fan Charts
- Iterated Forecasts

Review

- Optimal point forecast of y_{n+1} given information I_n is the conditional mean $E\left(y_{n+1}|I_n\right)$
- Estimate linear approximations by least-squares
- Combine point forecasts to reduce MSFE
- Select estimators and combination weights by cross-validation
- Estimate GARCH models for conditional variance

Interval Forecasts

- Take the form $[a, b]$
- Should contain y_{n+1} with probability $1 - 2\alpha$

$$
\begin{aligned}
1 - 2\alpha &= P_n\left(y_{n+1} \in [a, b]\right) \\
&= P_n\left(y_{n+1} \leq b\right) - P_n\left(y_{n+1} \leq a\right) \\
&= F_n(b) - F_n(a)
\end{aligned}
$$

  where $F_n(y)$ is the forecast distribution
- It follows that

$$
\begin{aligned}
a &= q_n(\alpha) \\
b &= q_n(1 - \alpha)
\end{aligned}
$$

- $a = \alpha$'th and $b = (1 - \alpha)$'th quantile of conditional distribution

Interval Forecasts are Conditional Quantiles

- The ideal 80% forecast interval, is the 10% and 90% quantile of the conditional distribution of y_{n+1} given I_n
- Our feasible forecast intervals are estimates of the 10% and 90% quantile of the conditional distribution of y_{n+1} given I_n
- The goal is to estimate conditional quantiles.

Mean-Variance Model

- Write

$$\begin{aligned} y_{t+1} &= \mu_t + \sigma_t \varepsilon_{t+1} \\ \mu_t &= E\left(y_{t+1} | I_t\right) \\ \sigma_t^2 &= \operatorname{var}\left(y_{t+1} | I_t\right) \end{aligned}$$

- Assume that ε_{t+1} is independent of I_t.
- Let $q_t(\alpha)$ and $q^{\varepsilon}(\alpha)$ be the α'th quantiles of y_{t+1} and ε_{t+1}. Then

$$q_t(\alpha) = \mu_t + \sigma_t q^{\varepsilon}(\alpha)$$

- Thus a $(1 - 2\alpha)$ forecast interval for y_{n+1} is

$$\left[\mu_n + \sigma_n q^{\varepsilon}(\alpha), \quad \mu_n + \sigma_n q^{\varepsilon}(1 - \alpha)\right]$$

Mean-Variance Model

- Given the conditional mean μ_n and variance σ_n^2, the conditional quantile of y_{n+1} is a linear function $\mu_n + \sigma_n q^{\varepsilon}(\alpha)$ of the conditional quantile $q^{\varepsilon}(\alpha)$ of the normalized error

$$\varepsilon_{n+1} = \frac{e_{n+1}}{\sigma_n}$$

- Interval forecasts thus can be summarized by μ_n, σ_n^2, and $q^{\varepsilon}(\alpha)$

Normal Error Quantile Forecasts

- Make the approximation $\varepsilon_{t+1} \sim N(0,1)$
  - Then $q^{\varepsilon}(\alpha) = Z(a)$ are normal quantiles
  - Useful simplification, especially in small samples
- 0.10, 0.25, 0.75, 0.90 quantiles are
  - -1.285, -0.675, 0.675, 1.285
- Forecast intervals

$$[\widehat{\mu}_n + \widehat{\sigma}_n Z(\alpha), \quad \widehat{\mu}_n + \widehat{\sigma}_n Z(1-\alpha)]$$

Nonparametric Error Quantile Forecasts

- Let $\varepsilon_{t+1} \sim F$ be unknown
  - We can estimate $q^{\varepsilon}(\alpha)$ as the empirical quantiles of the residuals
  - Set
$$\widehat{\varepsilon}_{t+1} = \frac{\widetilde{e}_{t+1}}{\widehat{\sigma}_t}$$
  - Sort $\widehat{\varepsilon}_1, ..., \widehat{\varepsilon}_n$.
  - $\widehat{q}^{\varepsilon}(\alpha)$ and $\widehat{q}^{\varepsilon}(1-\alpha)$ are the α'th and $(1-\alpha)$'th percentiles
$$[\widehat{\mu}_n + \widehat{\sigma}_n \widehat{q}^{\varepsilon}(\alpha), \quad \widehat{\mu}_n + \widehat{\sigma}_n \widehat{q}^{\varepsilon}(1-\alpha)]$$
- Computationally simple
- Reasonably accurate when $n \geq 100$
- Allows asymmetric and fat-tailed error distributions

Constant Variance Case

- If $\widehat{\sigma}_t = \widehat{\sigma}$ is a constant, there is no advantage for estimation of $\widehat{\sigma}$ for forecast interval
- Let $\widehat{q}^e(\alpha)$ and $\widehat{q}^e(1-\alpha)$ be the α'th and $(1-\alpha)$'th percentiles of original residuals $\widetilde{e}_{t+1}$
- Forecast Interval:

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

- When the estimated variance is a constant, this is numerically identical to the definition with rescaled errors $\widehat{\varepsilon}_{t+1}$

Computation in R

- *quadreg* package
  - may need to be installed
  - `library(quadreg)`
  - `rq` command
- If e is vector of (normalized) residuals and a is the quantile to be evalulated
  - `rq(e~1,a)`
  - `q=coef(rq(e~1,a))`
  - Quantile regression of e on an intercept

Example: Interest Rate Forecast

- $n = 603$ observations
- $\widehat{\varepsilon}_{t+1} = \dfrac{\widetilde{e}_{t+1}}{\widehat{\sigma}_t}$ from GARCH(1,1) model
- 0.10, 0.25, 0.75, 0.90 quantiles
- -1.16, -0.59, 0.62, 1.26
- Point Forecast = 1.96
- 50% Forecast interval = $[1.82, \quad 2.10]$
- 80% Forecast interval = $[1.69, \quad 2.25]$

Example: GDP

- $n = 207$ observations
- $\widehat{\varepsilon}_{t+1} = \dfrac{\widetilde{e}_{t+1}}{\widehat{\sigma}_t}$ from GARCH(1,1) model
- 0.10, 0.25, 0.75, 0.90 quantiles
- -1.18, -0.63, 0.57, 1.26
- Point Forecast $= 1.31$
- 50% Forecast interval $= [0.04, \quad 2.4]$
- 80% Forecast interval $= [-1.07, \quad 3.8]$

Mean-Variance Model Interval Forecasts - Summary

- The key is to break the distribution into the mean μ_t, variance σ_t^2 and the normalized error ε_{t+1}

$$y_{t+1} = \mu_t + \sigma_t \varepsilon_{t+1}$$

- Then the distribution of y_{n+1} is determined by μ_n, σ_n^2 and the distribution of ε_{n+1}
- Each of these three components can be separately approximated and estimated
- Typically, we put the most work into modeling (estimating) the mean μ_t
  - The remainder is modeled more simply
  - For macro forecasts, this reflects a belief (assumption?) that most of the predictability is in the mean, not the higher features.

Alternative Approach: Quantile Regression

- Recall, the ideal $1 - 2\alpha$ interval is $[q_n(\alpha), q_n(1-\alpha)]$
- $q_n(\alpha)$ is the α'th quantile of the one-step conditional distribution
- $F_n(y) = P\left(y_{n+1} \leq y \mid I_n\right)$
- Equivalently, let's directly model the conditional quantile function

Quantile Regression Function

- The conditional distribution is

$$P\left(y_{n+1} \leq y \mid I_n\right) \simeq P\left(y_{n+1} \leq y \mid \mathbf{x}_n\right)$$

- The conditional quantile function $q_\alpha(\mathbf{x})$ solves

$$P\left(y_{n+1} \leq q_\alpha(\mathbf{x}) \mid \mathbf{x}_n = \mathbf{x}\right) = \alpha$$

- $q_{.5}(\mathbf{x})$ is the conditional median
- $q_{.1}(\mathbf{x})$ is the 10% quantile function
- $q_{.9}(\mathbf{x})$ is the 90% quantile function

Quantile Regression Functions

- For each α, $q_\alpha(\mathbf{x})$ is an arbitrary function of $\mathbf{x}$
- For each $\mathbf{x}$, $q_\alpha(\mathbf{x})$ is monotonically increasing in α
- Quantiles are well defined even when moments are infinite
- When distributions are discrete then quantiles may be intervals – we ignore this
- We approximate the functions as linear in $q_\alpha(\mathbf{x})$

$$q_\alpha(\mathbf{x}) \simeq \mathbf{x}'\boldsymbol{\beta}_\alpha$$

  (after possible transformations in $\mathbf{x}$)
- The coefficient vector $\mathbf{x}'\boldsymbol{\beta}_\alpha$ depends on α

Linear Quantile Regression Functions

- $q_\alpha(\mathbf{x}) = \mathbf{x}'\boldsymbol{\beta}_\alpha$
- If only the intercept depends on α,

$$q_\alpha(\mathbf{x}) \simeq \mu_\alpha + \mathbf{x}'\boldsymbol{\beta}$$

  then the quantile regression lines are parallel
  - This is when the error e_{t+1} in a linear model is **independent** of the regressors
  - Strong conditional homoskedasticity
- In general, the coefficients are functions of α
  - Similar to conditional heteroskedasticity

Interval Forecasts

- An ideal $1-2\alpha$ interval forecast interval is

$$\left[\mathbf{x}_n'\boldsymbol{\beta}_\alpha, \quad \mathbf{x}_n'\boldsymbol{\beta}_{1-\alpha}\right]$$

- Note that the ideal point forecast is $\mathbf{x}_n'\boldsymbol{\beta}$ where $\boldsymbol{\beta}$ is the best linear predictor
- An alternative point forecast is the conditional median $\mathbf{x}_n'\boldsymbol{\beta}_{0.5}$
  - This has the property of being the best linear predictor in L_1 (mean absolute error)
- All are linear functions of $\mathbf{x}_n$, just different functions
- A feasible forecast interval is

$$\left[\mathbf{x}_n'\widehat{\boldsymbol{\beta}}_\alpha, \quad \mathbf{x}_n'\widehat{\boldsymbol{\beta}}_{1-\alpha}\right]$$

where $\widehat{\boldsymbol{\beta}}_\alpha$ and $\widehat{\boldsymbol{\beta}}_{1-\alpha}$ are estimates of $\boldsymbol{\beta}_\alpha$ and $\boldsymbol{\beta}_{1-\alpha}$

Check Function

- Recall that the mean $\mu = EY$ minimizes the L_2 risk $E(Y-m)^2$
- Similarly the median $q_{0.5}$ minimizes the L_1 risk $E|Y-m|$
- The α'th quantile q_α minimizes the "check function risk

$$E\rho_\alpha(Y-m)$$

where

$$\begin{aligned}\rho_\alpha(u) &= \begin{cases} -u(1-\alpha) & u<0 \\ u\alpha & u\geq 0 \end{cases} \\ &= u(\alpha - 1(u<0))\end{aligned}$$

- This is a tilted absolute value function
- To see the equivalence, evaluate the first order condition for minimization

Extremum Representation

- $q_\alpha(\mathbf{x})$ solves

$$q_\alpha(\mathbf{x}) = \underset{m}{\operatorname{argmin}}\, E\left(\rho_\alpha\left(y_{t+1} - m\right) | \mathbf{x}_t = \mathbf{x}\right)$$

- Sample criterion

$$S_\alpha(\boldsymbol{\beta}) = \frac{1}{n}\sum_{t=0}^{n-1} \rho_\alpha\left(y_{t+1} - \mathbf{x}_t'\boldsymbol{\beta}\right)$$

- Quantile regression estimator

$$\widehat{\boldsymbol{\beta}}_\alpha = \underset{\boldsymbol{\beta}}{\operatorname{argmin}}\, S_\alpha(\boldsymbol{\beta})$$

- Computation by linear programming
  - Stata
  - R
  - Matlab

Computation in R

- *quantreg* package
  - may need to be installed
  - `library(quantreg)`
  - For quantile regression of y on x at a'th quantile
    - do not include intercept in x, it will be automatically included
  - `rq(y~x,a)`
  - For coefficients,
    - `b=coef(rq(y~x,a))`

Distribution Theory

- The asymptotic theory for the dependent data case is not well developed
- The theory for the cross-section (iid) case is Angrist, Chernozhukov and Fernandez-Val (Econometrica, 2006)
- Their theory allows for quantile regression viewed as a best linear approximation

$$\sqrt{n}\left(\widehat{\beta}_{\alpha} - \beta_{\alpha}\right) \xrightarrow{d} N(0, V_{\alpha})$$

$$\begin{aligned} V_{\alpha} &= J_{\alpha}^{-1}\Sigma_{\alpha}J_{\alpha} \\ J_{\alpha} &= E\left(f_y\left(\mathbf{x}_t'\beta_{\alpha}|\mathbf{x}_t\right)\mathbf{x}_t\mathbf{x}_t'\right) \\ \Sigma_{\alpha} &= E\left(\mathbf{x}_t\mathbf{x}_t'u_t^2\right) \\ u_t &= 1\left(y_{t+1} < \mathbf{x}_t'\beta_{\alpha}\right) - \alpha \end{aligned}$$

- Under correct specification, $\Sigma_{\alpha} = \alpha(1-\alpha)E\left(\mathbf{x}_t\mathbf{x}_t'\right)$
- I suspect that this theorem extends to dependent data if the score is uncorrelated (dynamics are well specified)

Standard Errors

- The asymptotic variance depends on the conditional density function
  - Nonparametric estimation!
- To avoid this, most researchers use bootstrap methods
- For dependent data, this has not been explored
- Recommend: Use current software, but be cautious!

Crossing Problem and Solution

- The conditional quantile functions $q_\alpha(\mathbf{x})$ are monotonically increasing in α
- But the linear quantile regression approximations $q_\alpha(\mathbf{x}) \simeq \mathbf{x}'\boldsymbol{\beta}_\alpha$ cannot be globally monotonic in α, unless all lines are parallel
- The regression approximations may cross!
- The estimates $\widehat{q}_\alpha(\mathbf{x}) = \mathbf{x}'\widehat{\boldsymbol{\beta}}_\alpha$ may cross!
- If this happens, forecast intervals may be inverted:
  - A 90% interval may not nest an 80% interval
- Simple Solution: Reordering
  - If $\widehat{q}_{\alpha_1}(\mathbf{x}) > \widehat{q}_{\alpha_2}(\mathbf{x})$ when $\alpha_1 < \alpha_2 < \frac{1}{2}$, simply set $\widehat{q}_{\alpha_1}(\mathbf{x}) = \widehat{q}_{\alpha_2}(\mathbf{x})$, and conversely quantiles above $\frac{1}{2}$
  - Take the wider interval
  - Then the endpoint of the two intervals will be the same

Model Selection and Combination

- To my knowledge, no theory of model selection for median regression or quantile regression, even in iid context
- A natural conjecture is to use cross-validation on the sample check function
  - But no current theory justifies this choice
- My recommendation for model selection (or combination)
  - Select the model for the conditional mean by cross-validation
  - Use the same variables for all quantiles
  - Select the weights by cross-validation on the conditional mean
  - For each quantile, estimate the models with positive weights
  - Take the weighted combination using the same weights.

Example: Interest Rates

- AR(2) Specification (selected for regression by CV)

$$y_{t+1} = \beta_0 + \beta_1 y_t + \beta_2 y_{t-1} + e_t$$

	$\alpha = 0.10$	$\alpha = 0.25$	$\alpha = 0.75$	$\alpha = 0.90$
β_0	-0.31	-0.14	0.15	0.29
β_1	0.46	0.31	0.35	0.34
β_2	-0.22	-0.17	-0.21	-0.25

- Forecast 10% quantile

$$q_{0.1}(x_n) = -0.31 + 0.46 y_n - 0.22 y_{n-1}$$

- 50% Forecast interval $= [1.84, \quad 2.12]$
- 80% Forecast interval $= [1.65, \quad 2.25]$
- Very close to those from mean-variance estimates

Example: GDP

- Leading Indicator Model

$$y_{t+1} = \beta_0 + \beta_1 y_t + \beta_2 Spread_t + \beta_3 HighYield + \beta_4 Starts + \beta_5 Permits + e$$

	$\alpha = 0.10$	$\alpha = 0.25$	$\alpha = 0.75$	$\alpha = 0.90$
β_0	-2.72	-0.14	0.10	2.0
β_1	0.28	0.14	0.33	0.28
β_2	1.17	0.75	0.31	-0.14
β_3	-2.12	-1.83	0.62	0.37
β_4	-2.20	-0.44	6.68	11.4
β_5	3.45	1.61	-4.87	-9.53

- 50% Forecast interval $= [0.1, \quad 3.2]$
- 80% Forecast interval $= [-1.8, \quad 4.0]$

Distribution Forecasts

- The conditional distribution is

$$F_t(y) = P\left(y_{t+1} \leq y \mid I_t\right)$$

- It is not common to directly report $F_t(y)$
  - or the one-step forecast distribution $F_n(y)$
- However, $F_t(y)$ may be used as an input
- For example, simulation
- We thus may want an estimate $\widehat{F}_t(y)$ of $F_t(y)$

Mean-Variance Model Distribution Forecasts

- Model
$$y_{t+1} = \mu_t + \sigma_t \varepsilon_{t+1}$$
with ε_{t+1} is independent of I_t.
- Let ε_{t+1} have distribution $F^{\varepsilon}(u) = P(\varepsilon_t \leq u)$.
- The conditional distribution of y_{t+1} is
$$F_t(y) = F^{\varepsilon}\left(\frac{y_{t+1} - \mu_t}{\sigma_t}\right)$$
- Estimation
$$\widehat{F}_t(y) = \widehat{F}^{\varepsilon}\left(\frac{y_{t+1} - \widehat{\mu}_t}{\widehat{\sigma}_t}\right)$$
where $\widehat{F}^{\varepsilon}(u)$ is an estimate of $F^{\varepsilon}(u) = P(\varepsilon_t \leq u)$.

Normal Error Model

- Under the assumption $\varepsilon_{t+1} \sim N(0,1)$, $F^{\varepsilon}(u) = \Phi(u)$, the normal CDF
$$\widehat{F}_t(y) = \Phi\left(\frac{y - \widehat{\mu}_t}{\widehat{\sigma}_t}\right)$$
- To simulate from $\widehat{F}_t(y)$
  - Calculate $\widehat{\mu}_t$ and $\widehat{\sigma}_t$
  - Draw ε^*_{t+1} iid from $N(0,1)$
  - $y^*_{t+1} = \widehat{\mu}_t + \widehat{\sigma}_t \varepsilon^*_{t+1}$
- The normal assumption can be used when sample size n is very small
- But then $\widehat{F}_t(y)$ contains no information beyond $\widehat{\mu}_t$ and $\widehat{\sigma}_t$

Nonparametric Error Model

- Let $\widehat{F}_n^{\varepsilon}$ be the empirical distribution function (EDF) of the normalized residuals $\widehat{\varepsilon}_{t+1}$
- The EDF puts probability mass $1/n$ at each point $\{\widehat{\varepsilon}_1, ..., \widehat{\varepsilon}_n\}$

$$\widehat{F}_n^{\varepsilon}(u) = n^{-1} \sum_{t=0}^{n-1} 1\left(\widehat{\varepsilon}_{t+1} \leq u\right)$$

$$\begin{aligned}
\widehat{F}_t(y) &= \widehat{F}_n^{\varepsilon}\left(\frac{y - \widehat{\mu}_t}{\widehat{\sigma}_t}\right) \\
&= n^{-1} \sum_{j=0}^{n-1} 1\left(\frac{y - \widehat{\mu}_t}{\widehat{\sigma}_t} \leq \widehat{\varepsilon}_{j+1}\right) \\
&= n^{-1} \sum_{j=0}^{n-1} 1\left(y \leq \widehat{\mu}_t + \widehat{\sigma}_t \widehat{\varepsilon}_{j+1}\right)
\end{aligned}$$

- Notice the summation over j, holding $\widehat{\mu}_t, \widehat{\sigma}_t$ fixed

Simulate Estimated Conditional Distribution

- To simulate
  - Calculate $\widehat{\mu}_t$ and $\widehat{\sigma}_t$
  - Draw ε^*_{t+1} iid from normalized residuals $\{\widehat{\varepsilon}_1, ..., \widehat{\varepsilon}_n\}$
  - $y^*_{t+1} = \widehat{\mu}_t + \widehat{\sigma}_t \varepsilon^*_{t+1}$
  - y^*_{t+1} is a draw from $\widehat{F}_t(y)$

Plot Estimated Conditional Distribution

- $\widehat{F}_n(y) = n^{-1} \sum_{t=0}^{n-1} 1\left(y \leq \widehat{\mu}_n + \widehat{\sigma}_n \widehat{\varepsilon}_{t+1}\right)$
- A step function, with steps of height $1/n$ at $\widehat{\mu}_n + \widehat{\sigma}_n \widehat{\varepsilon}_{t+1}$
- Calculation
  - Calculate $\widehat{\mu}_n$, $\widehat{\sigma}_n$, and $y^*_{t+1} = \widehat{\mu}_n + \widehat{\sigma}_n \widehat{\varepsilon}_{t+1}$, $t = 0, ..., n-1$
  - Sort y^*_{t+1} into order statistics $y^*_{(j)}$
  - Equivalently, sort $\widehat{\varepsilon}_{t+1}$ into order statistics $\widehat{\varepsilon}_{(1)}$ and set $y^*_{(j)} = \widehat{\mu}_n + \widehat{\sigma}_n \widehat{\varepsilon}_{(j)}$
  - Plot on the y-axis $\{1/n, 2/n, 3/n, ..., 1\}$ against on the x-axis $y^*_{(1)}, y^*_{(2)}, ..., y^*_{(n)}$

Examples:

- Interest Rate
- GDP

Figure: 10-Year Bond Rate Forecast Distribution

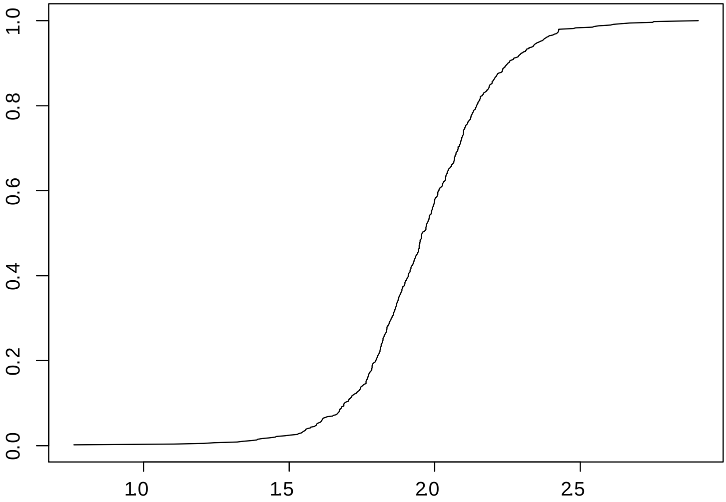
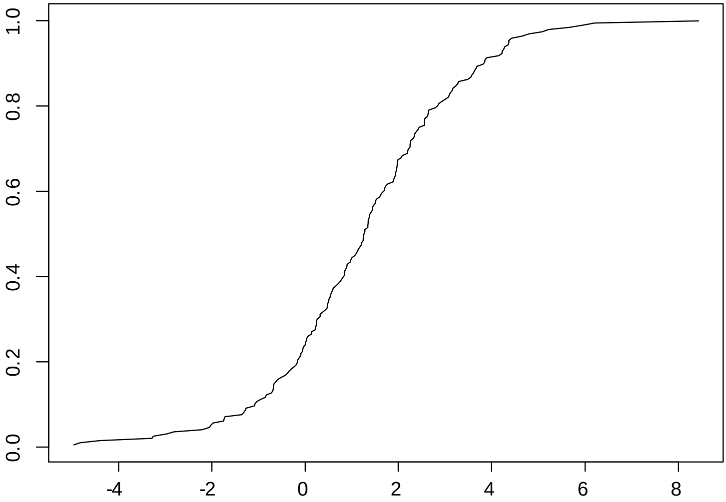


Figure: GDP Forecast Distribution



Quantile Regression Approach

- The distribution function may also be recovered from the estimated quantile functions.
- $F_n(q_\alpha(\mathbf{x}_n)) = \alpha$
- $\widehat{F}_n(\widehat{q}_\alpha(\mathbf{x}_n)) = \alpha$
- $\widehat{q}_\alpha(\mathbf{x}_n) = \mathbf{x}_n' \widehat{\boldsymbol{\beta}}_\alpha$
- Compute $\widehat{q}_\alpha(\mathbf{x}_n) = \mathbf{x}_n' \widehat{\boldsymbol{\beta}}_\alpha$ for a set of quantiles $\{\alpha_1, ..., \alpha_J\}$
- Plot α_j on the y-axis against $\widehat{q}_{\alpha_j}(\mathbf{x}_n)$ on the x-axis
  - The plot is $\widehat{F}_n(y)$ at $y = \widehat{q}_{\alpha_j}(\mathbf{x}_n)$
- If the quantile lines cross, then the plot will be non-monotonic
- The reordering method flattens the estimated distribution at these points

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

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

  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

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

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

Properties of the Error

- $E\left(\mathbf{x}_t e_{t+h}\right) = 0$
  - Projection
- $E\left(e_{t+h}\right) = 0$
  - Inclusion of an intercept
- The error e_{t+h} is NOT serially uncorrelated
- It is at least a MA(h-1)

Least Squares Estimation

$$
\begin{aligned}
\widehat{\beta} &= \left(\sum_{t=0}^{n-1} \mathbf{x}_t \mathbf{x}_t' \right)^{-1} \left(\sum_{t=0}^{n-1} \mathbf{x}_t y_{t+h} \right) \\
\widehat{y}_{n+h|n} &= \widehat{f}_{n+h|n} = \widehat{\beta}' \mathbf{x}_n
\end{aligned}
$$

Distribution Theory - Consistent Estimation

By the WLLN,

$$\begin{aligned}\widehat{\boldsymbol{\beta}} &= \left(\sum_{t=0}^{n-1} \mathbf{x}_t \mathbf{x}_t'\right)^{-1} \left(\sum_{t=0}^{n-1} \mathbf{x}_t y_{t+h}\right) \\ &\xrightarrow{p} \left(E\mathbf{x}_t \mathbf{x}_t'\right)^{-1} \left(E\mathbf{x}_t y_{t+h}\right) \\ &= \boldsymbol{\beta}\end{aligned}$$

Distribution Theory - Asymptotic Normality

By the dependent CLT,

$$\frac{1}{n}\sum_{t=0}^{n-1}\mathbf{x}_t e_{t+h} \xrightarrow{d} N(0,\Omega)$$

$$\begin{aligned}\Omega &= E\left(\mathbf{x}_t\mathbf{x}_t' e_{t+h}^2\right) + \sum_{j=1}^{\infty}\left(\mathbf{x}_t\mathbf{x}_{t+j}' e_{t+h}e_{t+h+j} + \mathbf{x}_{t+j}\mathbf{x}_t' e_{t+h}e_{t+h+j}\right)\\ &\simeq E\left(\mathbf{x}_t\mathbf{x}_t' e_{t+h}^2\right) + \sum_{j=1}^{h-1}\left(\mathbf{x}_t\mathbf{x}_{t+j}' e_{t+h}e_{t+h-j} + \mathbf{x}_{t+j}\mathbf{x}_t' e_{t+h}e_{t+h+j}\right)\end{aligned}$$

- A long-run (HAC) covariance matrix
- If model is correctly specified, the errors are a MA(h-1) and the sum truncates at $h-1$
- Otherwise, this is an approximation
- It does not simplify to the iid covariance matrix

Distribution Theory

- $\sqrt{n}\left(\widehat{\beta}-\beta\right) \xrightarrow{d} N(0, V)$
- $V = Q^{-1}\Omega Q^{-1}$
- $\Omega \approx E\left(\mathbf{x}_t\mathbf{x}_t' e_{t+h}^2\right) + \sum_{j=1}^{h-1}\left(\mathbf{x}_t\mathbf{x}_{t+j}' e_{t+h}e_{t+h-j} + \mathbf{x}_{t+j}\mathbf{x}_t' e_{t+h}e_{t+h+j}\right)$
- HAC variance matrix

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 $t+h$.

Algebraic Computation of Leave h out residuals

- Loop across each observation $t = (y_{t+h}, \mathbf{x}_t)$
- Leave out observations $t - h + 1, ..., t, ..., t + h - 1$
- R command
  - For positive integers i
  - x[-i] returns elements of x excluding indices i
  - Consider
    - `ii=seq(i-h+1,i+h-1)`
    - `ii<-ii[ii>0]`
    - `yi=y[-ii]`
    - `xi=x[-ii,]`
  - This removes $t - h + 1, ..., t, ..., t + h - 1$ from y and x

Variance Estimator

- Asymptotic variance (HAC) estimator with leave-h-out residuals

$$\begin{aligned}\widehat{V} &= \widehat{Q}^{-1}\widehat{\Omega}\widehat{Q}^{-1} \\ \widehat{Q} &= \frac{1}{n}\sum_{t=0}^{n-1}\mathbf{x}_t\mathbf{x}_t' \\ \widehat{\Omega} &= \frac{1}{n}\sum_{t=1}^{n}\mathbf{x}_t\mathbf{x}_t'\widetilde{e}_{t+h}^2 + \frac{1}{n}\sum_{j=1}^{h-1}\sum_{t=1}^{n-j}\left(\mathbf{x}_t\mathbf{x}_{t+j}'\widetilde{e}_{t+h}\widetilde{e}_{t+h+j} + \mathbf{x}_{t+j}\mathbf{x}_t'\widetilde{e}_{t+h}\widetilde{e}_{t+h+}\right.\end{aligned}$$

- Can use least-squares residuals $\widehat{e}_{t+h}$ instead of leave-h-out residuals, but then multiply $\widehat{V}$ by $n/(n - \dim(\mathbf{x}_t))$.
- Standard errors for $\widehat{\beta}$ are the square roots of the diagonal elements of $n^{-1}\widehat{V}$

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
2012:2	1.3
2012:3	1.6
2012:4	2.9
2013:1	2.2
2013:2	2.4
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

Point Forecast and MSFE

- Given an estimate $\widehat{\beta}(m)$ of β , the point forecast for y_{n+h} is

$$f_{n+h|n} = \widehat{\beta}' \mathbf{x}_n$$

- The mean-squared-forecast-error (MSFE) is

$$\begin{aligned} MSFE &= E\left(e_{n+h} - \mathbf{x}_n'\left(\widehat{\beta} - \beta\right)\right)^2 \\ &\simeq \sigma^2 + E\left(\left(\widehat{\beta} - \beta\right)' Q\left(\widehat{\beta} - \beta\right)\right) \end{aligned}$$

where $Q = E\left(\mathbf{x}_n \mathbf{x}_n'\right)$ and $\sigma^2 = E\left(e_{n+h}^2\right)$

- Same form as 1-step case

Residual Fit

$$\begin{aligned}\widehat{\sigma}^2 &= \frac{1}{n}\sum_{t=0}^{n-1} e_{t+h}^2 + \frac{1}{n}\sum_{t=0}^{n-1}\left(\mathbf{x}_t'\left(\widehat{\beta}-\beta\right)\right)^2 \\ &\quad -\frac{2}{n}\sum_{t=0}^{n-1} e_{t+h}\mathbf{x}_t'\left(\widehat{\beta}-\beta\right) \\ &\simeq MSFE - \frac{2}{n}\mathbf{e}'\mathbf{P}\mathbf{e}\end{aligned}$$

$$E\left(\widehat{\sigma}^2\right) \simeq MSFE_n - \frac{2}{n}B$$

where $B = E\left(\mathbf{e}'\mathbf{P}\mathbf{e}\right)$
Save form as 1-step case

Asymptotic Penalty

$$\mathbf{e}'\mathbf{P}\mathbf{e} = \left(\frac{1}{\sqrt{n}}\mathbf{e}'\mathbf{X}\right)\left(\frac{1}{n}\mathbf{X}'\mathbf{X}\right)^{-1}\left(\frac{1}{\sqrt{n}}\mathbf{X}'\mathbf{e}\right)$$
$$\longrightarrow_d Z'Q^{-1}Z$$

where $Z \sim N(0, \Omega)$, with $\Omega =$HAC variance.

$$\begin{aligned} B &= E\left(\mathbf{e}'\mathbf{P}\mathbf{e}\right) \\ &\longrightarrow \operatorname{tr}\left(Q^{-1}E\left(ZZ'\right)\right) \\ &= \operatorname{tr}\left(Q^{-1}\Omega\right) \end{aligned}$$

Ideal MSFE Criterion

$$C_n(m) = \widehat{\sigma}^2(m) + \frac{2}{n}\operatorname{tr}\left(Q^{-1}\Omega\right)$$

$$Q = E\left(\mathbf{x}_t \mathbf{x}_t'\right)$$

$$\Omega = E\left(\mathbf{x}_t \mathbf{x}_t' e_{t+h}^2\right) + \sum_{j=1}^{h-1}\left(\mathbf{x}_t \mathbf{x}_{t+j}' e_{t+h} e_{t+h-j} + \mathbf{x}_{t+j} \mathbf{x}_t' e_{t+h} e_{t+h+j}\right)$$

H-Step Robust Mallows Criterion

$$C_n(m) = \widehat{\sigma}^2(m) + \frac{2}{n}\,\mathrm{tr}\left(\widehat{Q}^{-1}\widehat{\Omega}\right)$$

where $\widehat{\Omega}$ is a HAC covariance matrix

H-Step Cross-Validation for Selection

$$CV_n(m) = \frac{1}{n} \sum_{i=0}^{n-1} \widetilde{e}_{t+h}(m)^2$$

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

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

Theorem: $E\left(CV_n(m)\right) \simeq MSFE(m)$
Thus $\widehat{m} = \operatorname{argmin} CV_n(m)$ is an estimate of $m = \operatorname{argmin} MSFE_n(m)$, but there is no proof of optimality

H-Step Cross-Validation for Forecast Combination

$$
\begin{aligned}
CV_n(\mathbf{w}) &= \frac{1}{n}\sum_{t=1}^{n} \widetilde{e}_{t+1}(\mathbf{w})^2 \\
&= \frac{1}{n}\sum_{t=1}^{n} \left(\sum_{m=1}^{M} w(m)\widetilde{e}_{t+1}(m) \right)^2 \\
&= \sum_{m=1}^{M}\sum_{\ell=1}^{M} w(m)w(\ell)\frac{1}{n}\sum_{t=1}^{n} \widetilde{e}_{t+1}(m)\widetilde{e}_{t+1}(\ell) \\
&= \mathbf{w}'\widetilde{\mathbf{S}}\mathbf{w}
\end{aligned}
$$

where

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

is covariance matrix of leave-h-out residuals.

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

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

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

Illustration 1

- $k = 8$ regressors
  - intercept
  - normal AR(1)'s with coefficient $\rho = 0.9$
- h-step error
  - normal MA(h-1)
  - equal coefficients
- Regression coefficients
  - $\beta = (\mu, 0, ..., 0)$
  - $n = 50$
  - MSPE plotted as a function of μ

Estimators

- Unconstrained Least-Squares
- Leave-1-out CV Selection
- Leave-h-out CV Selection
- Leave-1-out CV Combination
- Leave-h-out CV Combination

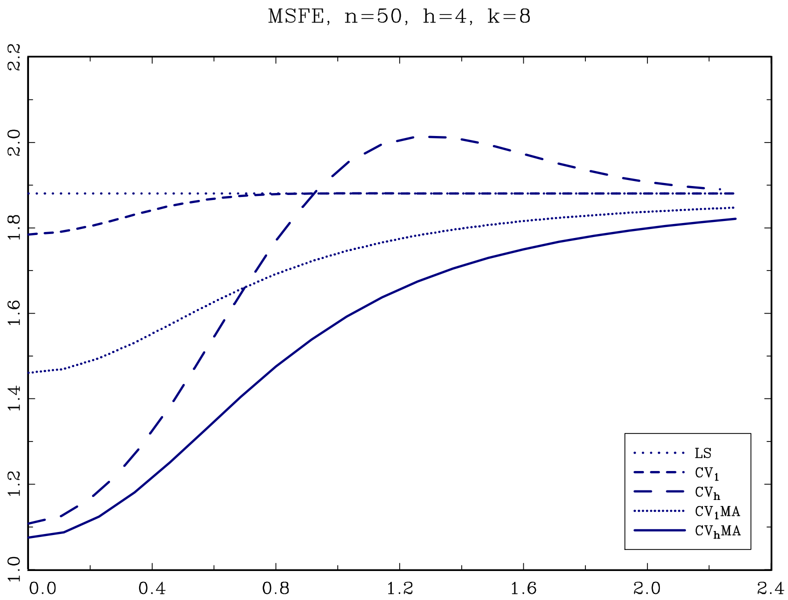

MSFE, n=50, h=4, k=8
1.0
1.2
1.4
1.6
1.8
2.0
2.2
0.0
0.4
0.8
1.2
1.6
2.0
2.4
LS
CV1
CVh
CV1MA
CVhMA

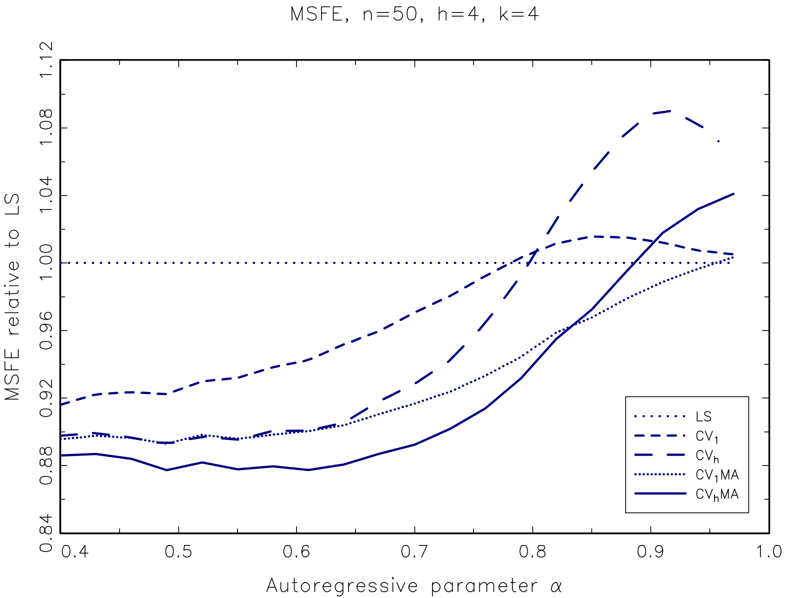
Illustration 2

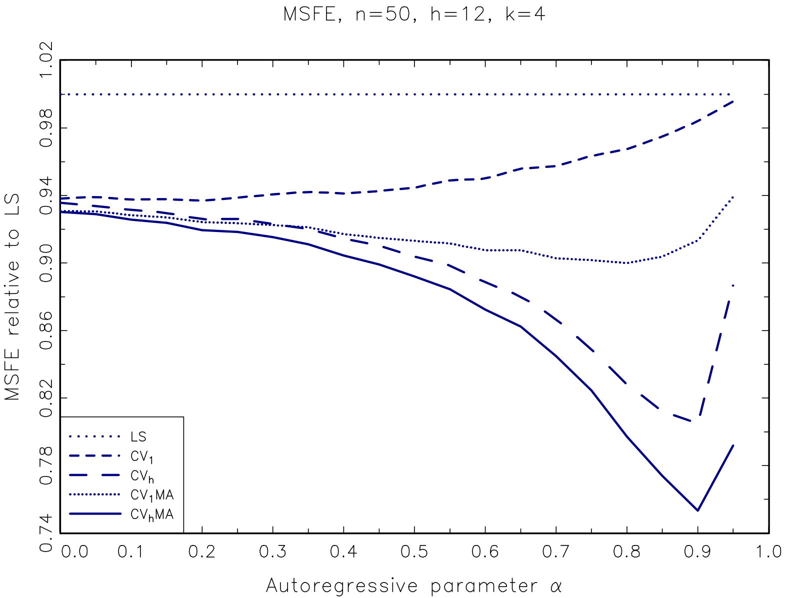
Model

$$y_t = \alpha y_{t-1} + u_t$$

Unconstrained model: AR(3)

$$y_t = \hat{\mu} + \hat{\beta}_1 y_{t-h} + \hat{\beta}_2 y_{t-h-1} + \hat{\beta}_3 y_{t-h-2} + \hat{e}_t$$


MSFE, n=50, h=4, k=4
MSFE relative to LS
1.12
1.08
1.04
1.00
0.96
0.92
0.88
0.84
0.4
0.5
0.6
0.7
0.8
0.9
1.0
Autoregressive parameter α
LS
CV_1
CV_h
CV_1MA
CV_hMA


MSFE, n=50, h=12, k=4
MSFE relative to LS
Autoregressive parameter α
1.02
0.98
0.94
0.90
0.86
0.82
0.78
0.74
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
LS
CV1
CVh
CV1MA
CVhMA

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						
$\widehat{y}_{n+h\mid n}$	1.7	2.0	1.9	2.0	2.1	2.3	2.6

h-step Variance Forecasting

- Not well developed using direct methods
- Suggest using constant variance specification

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:

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

Quantile Regression Approach

- $F_n(y) = P(y_{n+h} \leq y \mid I_n)$
- $q_\alpha(\mathbf{x}) \simeq \mathbf{x}'\boldsymbol{\beta}_\alpha$
- Estimate quantile regression of y_{t+h} on $\mathbf{x}_t$
- $1-2\alpha$ forecast interval is $[\mathbf{x}_n'\widehat{\boldsymbol{\beta}}_\alpha, \mathbf{x}_n'\widehat{\boldsymbol{\beta}}_{1-\alpha}]$
- Asymptotic theory not developed for $h-$step case
  - Developed for 1-step case
  - Extension is expected to work

Example: GDP Forecast Intervals (80%)

Using quantile regression approach

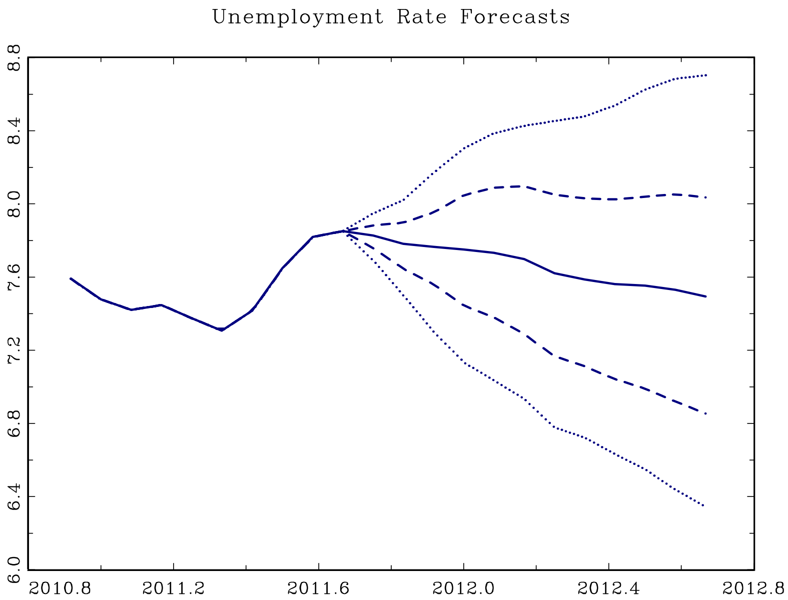
	$\widehat{y}_{n+h\|n}$	Interval
2012 : 2	1.3	$[-1.8, 4.1]$
2012 : 3	1.6	$[-0.4, 3.6]$
2012 : 4	2.0	$[-0.6, 4.6]$
2013 : 1	2.2	$[-0.3, 4.1]$
2013 : 2	2.4	$[0.2, 4.2]$
2013 : 3	2.7	$[0.6, 3.8]$
2013 : 4	2.9	$[0.7, 4.8]$
2014 : 1	3.2	$[1.5, 4.8]$

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)

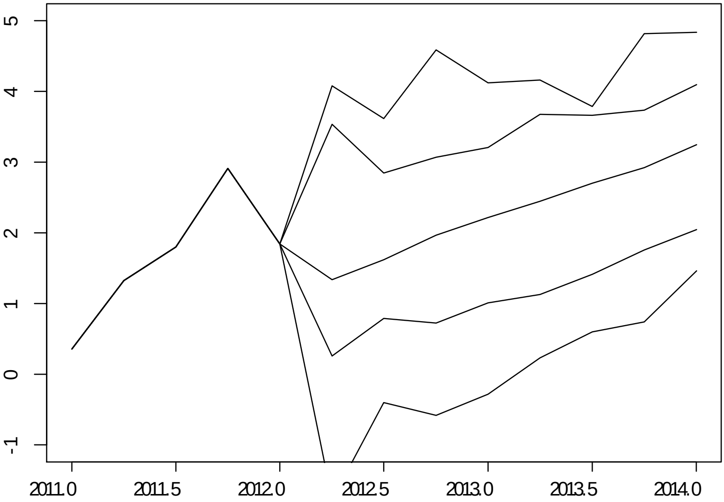

Unemployment Rate Forecasts
6.0
6.4
6.8
7.2
7.6
8.0
8.4
8.8
2010.8
2011.2
2011.6
2012.0
2012.4
2012.8

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

Iterated Forecasts

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

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

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 $y_{n+2}^* = \widehat{\mu}\left(\mathbf{x}_{n+1}^*\right) + \widehat{\sigma}\left(\mathbf{x}_{n+1}^*\right)\varepsilon_{t+2}^*$
  - 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

Assignment

- Take your favorite model from yesterday's assignment
- Calculate forecast intervals
- Make 1 through 12 step forecasts
  - point
  - interval
- Create a fan chart