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

Bruce E. Hansen

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
Optimal point forecast of $y_{n+1}$ given information $I_n$ is the conditional mean $E(y_{n+1} \mid I_n)$

- 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 \(1 - 2\alpha\)

\[
1 - 2\alpha = P_n (y_{n+1} \in [a, b]) \\
= P_n (y_{n+1} \leq b) - P_n (y_{n+1} \leq a) \\
= F_n(b) - F_n(a)
\]

where \(F_n(y)\) is the forecast distribution

- It follows that

\[
a = q_n(\alpha) \\
b = q_n(1 - \alpha)
\]

\(a = \alpha\)'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

\[
y_{t+1} = \mu_t + \sigma_t \varepsilon_{t+1}
\]
\[
\mu_t = E(y_{t+1}|l_t)
\]
\[
\sigma_t^2 = \text{var}(y_{t+1}|l_t)
\]

- Assume that \( \varepsilon_{t+1} \) is independent of \( l_t \).

- Let \( q_t(\alpha) \) and \( q^\varepsilon(\alpha) \) be the \( \alpha \)'th quantiles of \( y_{t+1} \) and \( \varepsilon_{t+1} \). Then

\[
q_t(\alpha) = \mu_t + \sigma_t q^\varepsilon(\alpha)
\]

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

\[
[\mu_n + \sigma_n q^\varepsilon(\alpha), \; \mu_n + \sigma_n q^\varepsilon(1 - \alpha)]
\]
Mean-Variance Model

- Given the conditional mean $\mu_n$ and variance $\sigma^2_n$, 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 $\mu_n$, $\sigma^2_n$, and $q^\varepsilon(\alpha)$
Normal Error Quantile Forecasts

- Make the approximation $\varepsilon_{t+1} \sim \mathcal{N}(0, 1)$
  - 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

$$[\hat{\mu}_n + \hat{\sigma}_n Z(\alpha), \hat{\mu}_n + \hat{\sigma}_n Z(1 - \alpha)]$$
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
    \[
    \hat{\varepsilon}_{t+1} = \frac{\tilde{e}_{t+1}}{\hat{\sigma}_t}
    \]
  - Sort \( \hat{\varepsilon}_1, ..., \hat{\varepsilon}_n \).
  - \( \hat{q}^\varepsilon(\alpha) \) and \( \hat{q}^\varepsilon(1 - \alpha) \) are the \( \alpha \)'th and \( (1 - \alpha) \)'th percentiles
    \[
    [\hat{\mu}_n + \hat{\sigma}_n \hat{q}^\varepsilon(\alpha), \hat{\mu}_n + \hat{\sigma}_n \hat{q}^\varepsilon(1 - \alpha)]
    \]
- Computationally simple
- Reasonably accurate when \( n \geq 100 \)
- Allows asymmetric and fat-tailed error distributions
Constant Variance Case

- If $\hat{\sigma}_t = \hat{\sigma}$ is a constant, there is no advantage for estimation of $\hat{\sigma}$ for forecast interval.
- Let $\hat{q}^e(\alpha)$ and $\hat{q}^e(1 - \alpha)$ be the $\alpha$'th and $(1 - \alpha)$'th percentiles of original residuals $\tilde{e}_{t+1}$.
- Forecast Interval:
  \[
  [\hat{\mu}_n + \hat{q}^e(\alpha), \quad \hat{\mu}_n + \hat{q}^e(1 - \alpha)]
  \]
- When the estimated variance is a constant, this is numerically identical to the definition with rescaled errors $\hat{\varepsilon}_{t+1}$. 

Bruce Hansen (University of Wisconsin)  Forecasting  July 23-27, 2012  10 / 102
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 evaluated
  - `rq(e~1,a)`
  - `q=coef(rq(e~1,a))`
  - Quantile regression of $e$ on an intercept
Example: Interest Rate Forecast

- \( n = 603 \) observations
- \( \hat{\varepsilon}_{t+1} = \frac{\tilde{e}_{t+1}}{\hat{\sigma}_t} \) 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 = \([1.82, 2.10]\)
- 80% Forecast interval = \([1.69, 2.25]\)
Example: GDP

- \( n = 207 \) observations
- \( \hat{\varepsilon}_{t+1} = \frac{\hat{e}_{t+1}}{\hat{\sigma}_t} \) 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 = \([0.04, 2.4]\)
- 80% Forecast interval = \([-1.07, 3.8]\)
The key is to break the distribution into the mean $\mu_t$, variance $\sigma^2_t$ and the normalized error $\varepsilon_{t+1}$

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

Then the distribution of $y_{n+1}$ is determined by $\mu_n$, $\sigma^2_n$ and the distribution of $\varepsilon_{n+1}$

Each of these three components can be separately approximated and estimated

Typically, we put the most work into modeling (estimating) the mean $\mu_t$

- 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 $\alpha$'th quantile of the one-step conditional distribution
- $F_n(y) = P(y_{n+1} \leq y \mid I_n)$
- Equivalently, let’s directly model the conditional quantile function
Quantile Regression Function

- The conditional distribution is
  \[ P(y_{n+1} \leq y \mid l_n) \simeq P(y_{n+1} \leq y \mid x_n) \]

- The conditional quantile function \( q_\alpha(x) \) solves
  \[ P(y_{n+1} \leq q_\alpha(x) \mid x_n = x) = \alpha \]

- \( q_{0.5}(x) \) is the conditional median
- \( q_{0.1}(x) \) is the 10\% quantile function
- \( q_{0.9}(x) \) is the 90\% quantile function
Quantile Regression Functions

- For each $\alpha$, $q_\alpha(x)$ is an arbitrary function of $x$
- For each $x$, $q_\alpha(x)$ is monotonically increasing in $\alpha$
- 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 $q_\alpha(x)$

$$q_\alpha(x) \simeq x' \beta_\alpha$$

(after possible transformations in $x$)

- The coefficient vector $x' \beta_\alpha$ depends on $\alpha$
Linear Quantile Regression Functions

- \( q_\alpha(x) = x' \beta_\alpha \)
- If only the intercept depends on \( \alpha \),

\[
q_\alpha(x) \sim \mu_\alpha + x' \beta
\]

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 \( \alpha \)
  - Similar to conditional heteroskedasticity
Interval Forecasts

- An ideal $1 - 2\alpha$ interval forecast interval is
  \[ [x'_n \beta_\alpha, x'_n \beta_{1-\alpha}] \]

- Note that the ideal point forecast is $x'_n \beta$ where $\beta$ is the best linear predictor.

- An alternative point forecast is the conditional median $x'_n \beta_{0.5}$
  - This has the property of being the best linear predictor in $L_1$ (mean absolute error).

- All are linear functions of $x_n$, just different functions.

- A feasible forecast interval is
  \[ [x'_n \hat{\beta}_\alpha, x'_n \hat{\beta}_{1-\alpha}] \]
  where $\hat{\beta}_\alpha$ and $\hat{\beta}_{1-\alpha}$ are estimates of $\beta_\alpha$ and $\beta_{1-\alpha}$.
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 $\alpha$'th quantile $q_\alpha$ minimizes the “check function risk

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

where

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

$$= u(\alpha - 1(u < 0))$$

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

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

$$q_{\alpha}(x) = \arg\min_{m} E \left( \rho_{\alpha} (y_{t+1} - m) \mid x_t = x \right)$$

- Sample criterion

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

- Quantile regression estimator

$$\hat{\beta}_{\alpha} = \arg\min_{\beta} S_{\alpha}(\beta)$$

- 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:

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

\[
V_\alpha = J_\alpha^{-1} \Sigma_\alpha J_\alpha
\]

\[
J_\alpha = E \left( f_y (x_t' \beta_\alpha | x_t) x_t x_t' \right)
\]

\[
\Sigma_\alpha = E \left( x_t x_t' u_t^2 \right)
\]

\[
u_t = 1 \left( y_{t+1} < x_t' \beta_\alpha \right) - \alpha
\]

- Under correct specification, \( \Sigma_\alpha = \alpha (1 - \alpha) E (x_t x_t') \)
- 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 $q_\alpha(x)$ are monotonically increasing in $\alpha$
- But the linear quantile regression approximations $q_\alpha(x) \approx x'\beta_\alpha$ cannot be globally monotonic in $\alpha$, unless all lines are parallel
- The regression approximations may cross!
- The estimates $\hat{q}_\alpha(x) = x'\hat{\beta}_\alpha$ may cross!
- If this happens, forecast intervals may be inverted:
  - A 90% interval may not nest an 80% interval

Simple Solution: Reordering

- If $\hat{q}_{\alpha_1}(x) > \hat{q}_{\alpha_2}(x)$ when $\alpha_1 < \alpha_2 < \frac{1}{2}$, simply set $\hat{q}_{\alpha_1}(x) = \hat{q}_{\alpha_2}(x)$, 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)**

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

\[
\begin{align*}
\alpha &= 0.10 & \alpha &= 0.25 & \alpha &= 0.75 & \alpha &= 0.90 \\
\beta_0 &= -0.31 & \beta_0 &= -0.14 & \beta_0 &= 0.15 & \beta_0 &= 0.29 \\
\beta_1 &= 0.46 & \beta_1 &= 0.31 & \beta_1 &= 0.35 & \beta_1 &= 0.34 \\
\beta_2 &= -0.22 & \beta_2 &= -0.17 & \beta_2 &= -0.21 & \beta_2 &= -0.25
\end{align*}
\]

- **Forecast 10% quantile**

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

- **50% Forecast interval** = \([1.84, \ 2.12]\)

- **80% Forecast interval** = \([1.65, \ 2.25]\)

- **Very close to those from mean-variance estimates**
Example: GDP

- Leading Indicator Model

\[ y_{t+1} = \beta_0 + \beta_1 y_t + \beta_2 \text{Spread}_t + \beta_3 \text{HighYield} + \beta_4 \text{Starts} + \beta_5 \text{Permits} + \epsilon_t \]

\[
\begin{array}{cccccc}
\alpha &= 0.10 & \alpha &= 0.25 & \alpha &= 0.75 & \alpha &= 0.90 \\
\beta_0 &= -2.72 & \beta_0 &= -0.14 & \beta_0 &= 0.10 & \beta_0 &= 2.0 \\
\beta_1 &= 0.28 & \beta_1 &= 0.14 & \beta_1 &= 0.33 & \beta_1 &= 0.28 \\
\beta_2 &= 1.17 & \beta_2 &= 0.75 & \beta_2 &= 0.31 & \beta_2 &= -0.14 \\
\beta_3 &= -2.12 & \beta_3 &= -1.83 & \beta_3 &= 0.62 & \beta_3 &= 0.37 \\
\beta_4 &= -2.20 & \beta_4 &= -0.44 & \beta_4 &= 6.68 & \beta_4 &= 11.4 \\
\beta_5 &= 3.45 & \beta_5 &= 1.61 & \beta_5 &= -4.87 & \beta_5 &= -9.53 \\
\end{array}
\]

- 50% Forecast interval = [0.1, 3.2]
- 80% Forecast interval = [−1.8, 4.0]
The conditional distribution is

\[ F_t(y) = P(y_{t+1} \leq y \mid I_t) \]

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 \( \hat{F}_t(y) \) of \( F_t(y) \)
Mean-Variance Model Distribution Forecasts

- **Model**

\[ y_{t+1} = \mu_t + \sigma_t \varepsilon_{t+1} \]

with \( \varepsilon_{t+1} \) is independent of \( I_t \).

- **Let** \( \varepsilon_{t+1} \) 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**

\[ \hat{F}_t(y) = \hat{F}^\varepsilon \left( \frac{y_{t+1} - \hat{\mu}_t}{\hat{\sigma}_t} \right) \]

where \( \hat{F}^\varepsilon(u) \) 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
  \[ \hat{F}_t(y) = \Phi \left( \frac{y - \hat{\mu}_t}{\hat{\sigma}_t} \right) \]

- To simulate from $\hat{F}_t(y)$
  - Calculate $\hat{\mu}_t$ and $\hat{\sigma}_t$
  - Draw $\varepsilon_{t+1}^*$ iid from $N(0, 1)$
  - $y_{t+1}^* = \hat{\mu}_t + \hat{\sigma}_t \varepsilon_{t+1}^*$

- The normal assumption can be used when sample size $n$ is very small
- But then $\hat{F}_t(y)$ contains no information beyond $\hat{\mu}_t$ and $\hat{\sigma}_t$
Nonparametric Error Model

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

\[
\hat{F}_n^\varepsilon(u) = n^{-1} \sum_{t=0}^{n-1} 1(\hat{\varepsilon}_{t+1} \leq u)
\]

\[
\hat{F}_t(y) = \hat{F}_n^\varepsilon\left(\frac{y - \hat{\mu}_t}{\hat{\sigma}_t}\right)
\]

\[
= n^{-1} \sum_{j=0}^{n-1} 1\left(\frac{y - \hat{\mu}_t}{\hat{\sigma}_t} \leq \hat{\varepsilon}_{j+1}\right)
\]

\[
= n^{-1} \sum_{j=0}^{n-1} 1\left(y \leq \hat{\mu}_t + \hat{\sigma}_t \hat{\varepsilon}_{j+1}\right)
\]

- Notice the summation over $j$, holding $\hat{\mu}_t, \hat{\sigma}_t$ fixed
Simulate Estimated Conditional Distribution

- To simulate
  - Calculate $\hat{\mu}_t$ and $\hat{\sigma}_t$
  - Draw $\varepsilon_{t+1}^*$ iid from normalized residuals $\{\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_n\}$
  - $y_{t+1}^* = \hat{\mu}_t + \hat{\sigma}_t \varepsilon_{t+1}^*$
  - $y_{t+1}^*$ is a draw from $\tilde{F}_t(y)$
Plot Estimated Conditional Distribution

- $\hat{F}_n(y) = n^{-1} \sum_{t=0}^{n-1} 1(y \leq \hat{\mu}_n + \hat{\sigma}_n \hat{\epsilon}_{t+1})$
- A step function, with steps of height $1/n$ at $\hat{\mu}_n + \hat{\sigma}_n \hat{\epsilon}_{t+1}$
- Calculation
  - Calculate $\hat{\mu}_n$, $\hat{\sigma}_n$, and $y^*_{t+1} = \hat{\mu}_n + \hat{\sigma}_n \hat{\epsilon}_{t+1}$, $t = 0, ..., n-1$
  - Sort $y^*_{t+1}$ into order statistics $y^*_j$
  - Equivalently, sort $\hat{\epsilon}_{t+1}$ into order statistics $\hat{\epsilon}(1)$ and set $y^*_j = \hat{\mu}_n + \hat{\sigma}_n \hat{\epsilon}(j)$
  - 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.

\[ F_n(q_\alpha(x_n)) = \alpha \]
\[ \hat{F}_n(\hat{q}_\alpha(x_n)) = \alpha \]
\[ \hat{q}_\alpha(x_n) = x'_n\hat{\beta}_\alpha \]

Compute \( \hat{q}_\alpha(x_n) = x'_n\hat{\beta}_\alpha \) for a set of quantiles \( \{\alpha_1, \ldots, \alpha_J\} \)

Plot \( \alpha_j \) on the y-axis against \( \hat{q}_{\alpha_j}(x_n) \) on the x-axis

- The plot is \( \hat{F}_n(y) \) at \( y = \hat{q}_{\alpha_j}(x_n) \)

- 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 $y_{n+h} \mid I_n \sim F_h(y_{n+h} \mid I_n)$
Point Forecast

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

\[
  f_{n+h|h} = \arg\min_f E \left( (y_{n+h} - f)^2 \mid I_n \right)
\]

\[
  = E (y_{n+h} \mid I_n)
\]

- 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

\[ y_{t+1} = \alpha (\alpha y_{t-1} + u_t) + u_{t+1} \]

\[ = \alpha^2 y_{t-1} + \alpha u_t + u_{t+1} \]

or

\[ y_{t+2} = \alpha^2 y_t + e_{t+2} \]

\[ u_{t+2} = \alpha u_{t+1} + u_{t+2} \]

- Repeat \( h \) times

\[ y_{t+h} = \alpha^h y_t + e_{t+h} \]

\[ e_{t+h} = u_{t+h} + \alpha u_{t+h-1} + \alpha^2 u_{t+h-2} + \cdots + \alpha^{h-1} u_{t+1} \]
AR(1)

$h$-step forecast

\[ y_{t+h} = \alpha^h y_t + e_{t+h} \]
\[ e_{t+h} = u_{t+h} + \alpha u_{t+h-1} + \alpha^2 u_{t+h-2} + \cdots + \alpha^{h-1} u_{t+1} \]
\[ E \left( y_{n+h} \mid I_n \right) = \alpha^h y_n \]

- $h$-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

\[
E(y_{t+2}|I_t) = \alpha_0 + \alpha_1 E(y_{t+1}|I_t) + \alpha_2 E(y_t|I_t) + E(e_{t+2}|I_t)
\]
\[
= \alpha_0 + \alpha_1 (\alpha_0 + \alpha_1 y_t + \alpha_2 y_{t-1}) + \alpha_2 y_t
\]
\[
= \alpha_0 + \alpha_1 \alpha_0 + (\alpha_1^2 + \alpha_2) y_t + \alpha_1 \alpha_2 y_{t-1}
\]

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

\[ y_{t+1} = x_t' \beta + u_t \]

then

\[ E(y_{t+h}|l_t) = E(x_{t+h-1}|l_t)' \beta \]

If \( E(x_{t+h-1}|l_t) \) is itself (approximately) a linear function of \( x_t \), then

\[ E(y_{t+h}|l_t) = x_t' \gamma \]

\[ y_{t+h} = x_t' \gamma + e_{t+h} \]

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: $\Delta y_{t+h}$
  - Cumulative Change: $\Delta y_{t+h}$
    - Cumulative GDP growth
Forecast Transformation

- $f_{n+h|n} = E(y_{n+h}|I_n) = \text{expected future level}$
  - Level specification
    \[ y_{t+h} = x_t' \beta + e_{t+h} \]
    \[ f_{n+h|n} = x_t' \beta \]
  - Difference specification
    \[ \Delta y_{t+h} = x_t' \beta_h + e_{t+h} \]
    \[ f_{n+h|n} = y_n + x_t' \beta_1 + \cdots + x_t' \beta_h \]
  - Multi-Step difference specification
    \[ y_{t+h} - y_t = x_t' \beta + e_{t+h} \]
    \[ f_{n+h|n} = y_n + x_t' \beta \]
There are two methods of multistep ($h > 1$) forecasts

- **Direct Forecast**
  - Model and estimate $E(y_{n+h}|l_n)$ directly

- **Iterated Forecast**
  - Model and estimate one-step $E(y_{n+1}|l_n)$
  - Iterate forward $h$ steps
  - Requires full model for all variables

Both have advantages and disadvantages

- For now, we will focus on direct method.
Direct Multi-Step Forecasting

- **Markov approximation**
  
  \[ E (y_{n+h} | I_n) = E (y_{n+h} | x_n, x_{n-1}, ...) \approx E (y_{n+h} | x_n, ..., x_{n-p}) \]

- **Linear approximation**
  
  \[ E (y_{n+h} | x_n, ..., x_{n-p}) \approx \beta' x_n \]

- **Projection Definition**
  
  \[ \beta = (E (x_t x_t'))^{-1} (E (x_t y_{t+h})) \]

- **Forecast error**
  
  \[ e_{t+h} = y_{t+h} - \beta' x_t \]
Multi-Step Forecast Model

\[ y_{t+h} = \beta' x_t + e_{t+h} \]

\[ \beta = \left( E(x_t x'_t) \right)^{-1} E(x_t y_{t+h}) \]

\[ E(x_t e_{t+h}) = 0 \]

\[ \sigma^2 = E(e_{t+h}^2) \]
Properties of the Error

- \( E(x_t e_{t+h}) = 0 \)
  - Projection
- \( E(e_{t+h}) = 0 \)
  - Inclusion of an intercept
- The error \( e_{t+h} \) is NOT serially uncorrelated
- It is at least a MA(h-1)
Least Squares Estimation

\[ \hat{\beta} = \left( \sum_{t=0}^{n-1} x_t x_t' \right)^{-1} \left( \sum_{t=0}^{n-1} x_t y_{t+h} \right) \]

\[ \hat{y}_{n+h|n} = \hat{f}_{n+h|n} = \hat{\beta}' x_n \]
By the WLLN,

\[ \hat{\beta} = \left( \sum_{t=0}^{n-1} x_t x'_t \right)^{-1} \left( \sum_{t=0}^{n-1} x_t y_{t+h} \right) \]

\[ \xrightarrow{p} \left( E x_t x'_t \right)^{-1} \left( E x_t y_{t+h} \right) \]

\[ = \beta \]
Distribution Theory - Asymptotic Normality

By the dependent CLT,

\[ \frac{1}{n} \sum_{t=0}^{n-1} x_t e_{t+h} \xrightarrow{d} N(0, \Omega) \]

\[ \Omega = E(x_t x_t' e_{t+h}^2) + \sum_{j=1}^{\infty} (x_t x_{t+j}' e_{t+h} e_{t+h+j} + x_{t+j} x_t' e_{t+h} e_{t+h+j}) \]

\[ \simeq E(x_t x_t' e_{t+h}^2) + \sum_{j=1}^{h-1} (x_t x_{t+j}' e_{t+h} e_{t+h-j} + x_{t+j} x_t' e_{t+h} e_{t+h+j}) \]

- 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
\[ \sqrt{n} \left( \hat{\beta} - \beta \right) \xrightarrow{d} N(0, V) \]
\[ V = Q^{-1} \Omega Q^{-1} \]
\[ \Omega \approx E \left( x_t x_t' e_{t+h}^2 \right) + \sum_{j=1}^{h-1} \left( x_t x_{t+j}' e_{t+h} e_{t+h-j} + x_{t+j} x_t' e_{t+h} e_{t+h+j} \right) \]
HAC variance matrix
Residuals

- Least-squares residuals
  - $\tilde{e}_{t+h} = y_{t+h} - \hat{\beta}' x_t$
  - Standard, but overfit

- Leave-one-out residuals
  - $\tilde{e}_{t+h} = y_{t+h} - \hat{\beta}_{-t}' x_t$
  - Does not correct for MA errors

- Leave $h$ out residuals

$$\tilde{e}_{t+h} = y_{t+h} - \hat{\beta}'_{-t,h} x_t$$

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

- The summation is over all observations outside $h - 1$ periods of $t + h$. 
Algebraic Computation of Leave h out residuals

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

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

\[
\hat{V} = \hat{Q}^{-1}\hat{\Omega}\hat{Q}^{-1}
\]

\[
\hat{Q} = \frac{1}{n} \sum_{t=0}^{n-1} x_t x'_t
\]

\[
\hat{\Omega} = \frac{1}{n} \sum_{t=1}^{n} x_t x'_t \hat{e}^2_{t+h} + \frac{1}{n} \sum_{j=1}^{h-1} \sum_{t=1}^{n-j} (x_t x'_{t+j} \hat{e}_{t+h} \hat{e}_{t+h+j} + x_{t+j} x'_t \hat{e}_{t+h} \hat{e}_{t+h+j})
\]

- Can use least-squares residuals \( \hat{e}_{t+h} \) instead of leave-h-out residuals, but then multiply \( \hat{V} \) by \( n/(n - \text{dim}(x_t)) \).

- Standard errors for \( \hat{\beta} \) are the square roots of the diagonal elements of \( n^{-1}\hat{V} \).
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

<table>
<thead>
<tr>
<th>( h )</th>
<th>( \beta_0 )</th>
<th>( \Delta y_t )</th>
<th>( \Delta y_{t-1} )</th>
<th>( Spread_t )</th>
<th>( HighYield_t )</th>
<th>( HS_t )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.33 (1.0)</td>
<td>0.16 (.10)</td>
<td>0.09 (.10)</td>
<td>0.61 (.23)</td>
<td>-1.10 (.75)</td>
<td>1.86 (.65)</td>
</tr>
<tr>
<td>2</td>
<td>-0.38 (1.3)</td>
<td>0.18 (.09)</td>
<td>0.04 (.05)</td>
<td>0.65 (.19)</td>
<td>-0.68 (.70)</td>
<td>1.64 (.70)</td>
</tr>
<tr>
<td>3</td>
<td>-0.01 (1.6)</td>
<td>0.13 (.08)</td>
<td>0.05 (.07)</td>
<td>0.65 (.22)</td>
<td>-0.48 (.90)</td>
<td>1.31 (.80)</td>
</tr>
<tr>
<td>4</td>
<td>0.47 (1.8)</td>
<td>0.13 (.09)</td>
<td>0.02 (.06)</td>
<td>0.65 (.25)</td>
<td>-0.41 (1.01)</td>
<td>1.01 (.94)</td>
</tr>
</tbody>
</table>
Example: GDP Forecast

<table>
<thead>
<tr>
<th>Year</th>
<th>Cumulative Annualized Growth</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012:2</td>
<td>1.3</td>
</tr>
<tr>
<td>2012:3</td>
<td>1.6</td>
</tr>
<tr>
<td>2012:4</td>
<td>2.9</td>
</tr>
<tr>
<td>2013:1</td>
<td>2.2</td>
</tr>
<tr>
<td>2013:2</td>
<td>2.4</td>
</tr>
<tr>
<td>2013:3</td>
<td>2.7</td>
</tr>
<tr>
<td>2013:4</td>
<td>2.9</td>
</tr>
<tr>
<td>2014:1</td>
<td>3.2</td>
</tr>
</tbody>
</table>
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 \( \hat{\beta}(m) \) of \( \beta \), the point forecast for \( y_{n+h} \) is
  \[
  f_{n+h|n} = \hat{\beta}' x_n
  \]

- The mean-squared-forecast-error (MSFE) is
  \[
  MSFE = E \left( e_{n+h} - x_n' \left( \hat{\beta} - \beta \right) \right)^2
  \sim \sigma^2 + E \left( \left( \hat{\beta} - \beta \right)' Q \left( \hat{\beta} - \beta \right) \right)
  \]
  where \( Q = E \left( x_n x_n' \right) \) and \( \sigma^2 = E \left( e_{n+h}^2 \right) \)

- Same form as 1-step case
Residual Fit

\[ \hat{\sigma}^2 = \frac{1}{n} \sum_{t=0}^{n-1} e_{t+h}^2 + \frac{1}{n} \sum_{t=0}^{n-1} \left( x_t' \left( \hat{\beta} - \beta \right) \right)^2 - \frac{2}{n} \sum_{t=0}^{n-1} e_{t+h} x_t' \left( \hat{\beta} - \beta \right) \]

\[ \simeq MSFE - \frac{2}{n} e' P e \]

\[ E \left( \hat{\sigma}^2 \right) \simeq MSFE_n - \frac{2}{n} B \]

where \( B = E( e' P e ) \)

Save form as 1-step case
Asymptotic Penalty

\[ e' Pe = \left( \frac{1}{\sqrt{n}} e' X \right) \left( \frac{1}{n} X' X \right)^{-1} \left( \frac{1}{\sqrt{n}} X' e \right) \]

\[ \rightarrow_d Z' Q^{-1} Z \]

where \( Z \sim N(0, \Omega) \), with \( \Omega = \text{HAC variance} \).

\[ B = E (e' Pe) \]

\[ \rightarrow \text{tr} \left( Q^{-1} E (ZZ') \right) \]

\[ = \text{tr} \left( Q^{-1} \Omega \right) \]
Ideal MSFE Criterion

\[ C_n(m) = \hat{\sigma}^2(m) + \frac{2}{n} \text{tr} \left( Q^{-1} \Omega \right) \]

\[ Q = E \left( x_t x_t' \right) \]

\[ \Omega = E \left( x_t x_t' e_{t+h}^2 \right) + \sum_{j=1}^{h-1} \left( x_t x_t' e_{t+h} e_{t+h-j} + x_{t+j} x_t' e_{t+h} e_{t+h+j} \right) \]
H-Step Robust Mallows Criterion

\[ C_n(m) = \hat{\sigma}^2(m) + \frac{2}{n} \text{tr} \left( \hat{Q}^{-1} \hat{\Omega} \right) \]

where \( \hat{\Omega} \) is a HAC covariance matrix.
H-Step Cross-Validation for Selection

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

\[ \tilde{e}_{t+h} = y_{t+h} - \hat{\beta}_{-t,h}'x_t \]

\[ \hat{\beta}_{-t,h} = \left( \sum_{|j+h-t| \geq h} x_j x_j' \right)^{-1} \left( \sum_{|j+h-t| \geq h} x_j y_{j+h} \right) \]

**Theorem:** \( E(CV_n(m)) \sim MSFE(m) \)

Thus \( \hat{m} = \text{argmin} \ CV_n(m) \) is an estimate of \( m = \text{argmin} \ MSFE_n(m) \), but there is no proof of optimality.
H-Step Cross-Validation for Forecast Combination

\[ CV_n(w) = \frac{1}{n} \sum_{t=1}^{n} \tilde{e}_{t+1}(w)^2 \]

\[ = \frac{1}{n} \sum_{t=1}^{n} \left( \sum_{m=1}^{M} w(m) \tilde{e}_{t+1}(m) \right)^2 \]

\[ = \sum_{m=1}^{M} \sum_{\ell=1}^{M} w(m) w(\ell) \frac{1}{n} \sum_{t=1}^{n} \tilde{e}_{t+1}(m) \tilde{e}_{t+1}(\ell) \]

\[ = w' \tilde{S} w \]

where

\[ \tilde{S} = \frac{1}{n} \tilde{e}' \tilde{e} \]

is covariance matrix of leave-h-out residuals.
Cross-validation Weights

Combination weights found by constrained minimization of $CV_n(w)$

$$\min_w CV_n(w) = w' \tilde{S} w$$

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
  - $\beta = (\mu, 0, \ldots, 0)$
  - $n = 50$
  - MSPE plotted as a function of $\mu$
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
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

Autoregressive parameter $\alpha$

MSFE relative to LS

Legend:
- LS
- $CV_1$
- $CV_h$
- $CV_{1,MA}$
- $CV_{h,MA}$
MSFE, n=50, h=12, k=4

Autoregressive parameter $\alpha$
### Example: GDP Forecast Weights by Horizon

<table>
<thead>
<tr>
<th></th>
<th>$h = 1$</th>
<th>$h = 2$</th>
<th>$h = 3$</th>
<th>$h = 4$</th>
<th>$h = 5$</th>
<th>$h = 6$</th>
<th>$h = 7$</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR(1)</td>
<td>.15</td>
<td>.19</td>
<td>.28</td>
<td>.18</td>
<td>.16</td>
<td>.11</td>
<td></td>
</tr>
<tr>
<td>AR(2)</td>
<td>.30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AR(1)+HS</td>
<td>.66</td>
<td>.70</td>
<td>.22</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AR(1)+HS+BP</td>
<td></td>
<td>.14</td>
<td>.58</td>
<td>.72</td>
<td>.82</td>
<td>.84</td>
<td>.89</td>
</tr>
<tr>
<td>AR(2)+HS</td>
<td>.04</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$\hat{y}_{n+h|n}$

<p>| | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.7</td>
<td>2.0</td>
<td>1.9</td>
<td>2.0</td>
<td>2.1</td>
<td>2.3</td>
<td>2.6</td>
</tr>
</tbody>
</table>
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 \( \hat{q}^e(\alpha) \) and \( \hat{q}^e(1 - \alpha) \) be the \( \alpha \)'th and \( (1 - \alpha) \)'th percentiles of residuals \( \tilde{e}_{t+h} \)
- Forecast Interval:
  \[
  [\hat{\mu}_n + \hat{q}^e(\alpha), \quad \hat{\mu}_n + \hat{q}^e(1 - \alpha)]
  \]
Quantile Regression Approach

- \( F_n(y) = P(y_{n+h} \leq y \mid l_n) \)
- \( q_\alpha(x) \simeq x' \beta_\alpha \)
- Estimate quantile regression of \( y_{t+h} \) on \( x_t \)
- \( 1 - 2\alpha \) forecast interval is \([x'_n \hat{\beta}_\alpha, x'_n \hat{\beta}_{1-\alpha}]\)
- 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

| Year : Quarter | \( \hat{y}_{n+h|n} \) | 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, $h = 1, \ldots, H$
  - 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 $h = 1, \ldots, 12$ (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

The graph shows the forecasted unemployment rate over time from 2010.8 to 2012.8. The lines represent different forecasts, with the solid blue line indicating the most likely scenario. The data suggests a slight increase in the unemployment rate starting from 2011.2 and peaking around 2011.6, followed by a decrease towards 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

$$E(y_{t+1}|I_t) = A_0 + A_1 E(y_t|I_t) + \cdots + A_k E(y_{t-k+1}|I_t)$$

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

- 2-step conditional mean

$$E(y_{t+1}|I_{t-1}) = E(E(y_{t+1}|I_t)|I_{t-1})$$

$$= A_0 + A_1 E(y_t|I_{t-1}) + \cdots + A_k E(y_{t-k+1}|I_{t-1})$$

$$= A_0 + A_1 E(y_t|I_{t-1}) + A_2 y_{t-1} + \cdots + A_k y_{t-k+1}$$

- $h$-step conditional mean

$$E(y_{t+1}|I_{t-h+1}) = E(E(y_{t+1}|I_t)|I_{t-h+1})$$

$$= A_0 + A_1 E(y_t|I_{t-h+1}) + \cdots + A_k E(y_{t-k+1}|I_{t-h+1})$$

- Linear in lower-order (up to $h - 1$ step) conditional means
Iterative Least Squares Forecasts

- Estimate 1-step VAR(k) by least-squares
  \[ y_{t+1} = \hat{A}_0 + \hat{A}_1 y_t + \hat{A}_2 y_{t-1} + \cdots + \hat{A}_k y_{t-k+1} + \hat{u}_{t+1} \]

- Gives 1-step point forecast
  \[ \hat{y}_{n+1|n} = \hat{A}_0 + \hat{A}_1 y_n + \hat{A}_2 y_{n-1} + \cdots + \hat{A}_k y_{n-k+1} \]

- 2-step iterative forecast
  \[ \hat{y}_{n+2|n} = \hat{A}_0 + \hat{A}_1 \hat{y}_{n+1|n} + \hat{A}_2 y_n + \cdots + \hat{A}_k y_{n-k+2} \]

- \( h \)-step iterative forecast
  \[ \hat{y}_{n+h|n} = \hat{A}_0 + \hat{A}_1 \hat{y}_{n+h-1|n} + \hat{A}_2 \hat{y}_{n+h-2|n} + \cdots + \hat{A}_k \hat{y}_{n+h-k|n} \]

- This is (numerically) different than the direct LS forecast
Illustration 1: GDP Growth

- AR(2) Model
  \[ y_{t+1} = 1.6 + 0.30y_t + 0.16y_{t-1} \]
  \[ y_n = 1.8, \ y_{n-1} = 2.9 \]
  \[ \hat{y}_{n+1} = 1.6 + 0.30 \times 1.8 + 0.16 \times 2.9 = 2.6 \]
  \[ \hat{y}_{n+2} = 1.6 + 0.30 \times 2.6 + 0.16 \times 1.8 = 2.7 \]
  \[ \hat{y}_{n+3} = 1.6 + 0.30 \times 2.7 + 0.16 \times 2.6 = 2.9 \]
  \[ \hat{y}_{n+4} = 1.6 + 0.30 \times 2.9 + 0.16 \times 2.7 = 3.0 \]
### Point Forecasts

<table>
<thead>
<tr>
<th>Year</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012:2</td>
<td>2.65</td>
</tr>
<tr>
<td>2012:3</td>
<td>2.72</td>
</tr>
<tr>
<td>2012:4</td>
<td>2.87</td>
</tr>
<tr>
<td>2013:1</td>
<td>2.93</td>
</tr>
<tr>
<td>2013:2</td>
<td>2.97</td>
</tr>
<tr>
<td>2013:3</td>
<td>2.99</td>
</tr>
<tr>
<td>2013:4</td>
<td>3.00</td>
</tr>
<tr>
<td>2014:1</td>
<td>3.01</td>
</tr>
</tbody>
</table>
Illustration 2: GDP Growth + Housing Starts

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

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

<table>
<thead>
<tr>
<th>Year</th>
<th>GDP</th>
<th>Housing</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012:2</td>
<td>2.36</td>
<td>0.76</td>
</tr>
<tr>
<td>2012:3</td>
<td>2.38</td>
<td>0.80</td>
</tr>
<tr>
<td>2012:4</td>
<td>2.53</td>
<td>0.84</td>
</tr>
<tr>
<td>2013:1</td>
<td>2.58</td>
<td>0.88</td>
</tr>
<tr>
<td>2013:2</td>
<td>2.64</td>
<td>0.92</td>
</tr>
<tr>
<td>2013:3</td>
<td>2.66</td>
<td>0.95</td>
</tr>
<tr>
<td>2013:4</td>
<td>2.69</td>
<td>0.98</td>
</tr>
<tr>
<td>2014:1</td>
<td>2.71</td>
<td>1.01</td>
</tr>
</tbody>
</table>
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
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 \( \mu (x) \) and \( \sigma (x) \) denote the models for the conditional one-step mean and standard deviation as a function of the conditional variables \( x \).
- Let \( \hat{\mu} (x) \) and \( \hat{\sigma} (x) \) denote the estimates of these functions, and let \( \{\hat{\varepsilon}_1, ..., \hat{\varepsilon}_n\} \) be the normalized residuals.
- \( x_n = (y_n, y_{n-1}, ..., y_{n-p}) \) is known. Set \( x_n^* = x_n \).
- To create one \( h \)-step realization:
  - Draw \( \varepsilon_{n+1}^* \) iid from normalized residuals \( \{\hat{\varepsilon}_1, ..., \hat{\varepsilon}_n\} \).
  - Set \( y_{n+1}^* = \hat{\mu} (x_n^*) + \hat{\sigma} (x_n^*) \varepsilon_{t+1}^* \).
  - Set \( x_{n+1}^* = (y_{n+1}^*, y_n, ..., y_{n-p+1}) \).
  - Draw \( \varepsilon_{n+2}^* \) iid from normalized residuals \( \{\hat{\varepsilon}_1, ..., \hat{\varepsilon}_n\} \).
  - Set \( y_{n+2}^* = \hat{\mu} (x_{n+1}^*) + \hat{\sigma} (x_{n+1}^*) \varepsilon_{t+2}^* \).
  - Set \( x_{n+2}^* = (y_{n+2}^*, y_{n+1}^*, ..., y_{n-p+2}) \).
  - 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 $y^*_{n+h}(b)$, $b = 1, \ldots, B$
- 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 $h = 1, \ldots, H$
- 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