Forecasting
Lecture 1: Foundations

Bruce E. Hansen

Central Bank of Chile
October 29-31, 2013
3-Day Course

- **Tuesday**
  - Point Forecasting
  - Forecast Selection
  - Leading Indicators
  - Variance Forecasting
  - Interval Forecasting

- **Wednesday**
  - Combination Forecasts
  - Multi-Step Forecasting
  - Fan Charts

- **Thursday: Structural Breaks**
Course Website

- [www.ssc.wisc.edu/~bhansen/cbc](http://www.ssc.wisc.edu/~bhansen/cbc)
- Slides for all lectures
- Data for the lectures
- R code for empirical analysis for lectures 1 & 2
- Related: [www.ssc.wisc.edu/~bhansen/crete](http://www.ssc.wisc.edu/~bhansen/crete)
  - 5-day forecasting course
  - All the empirical analysis reported here was done in spring 2012, so the “forecasts” appear to be about the past
  - We can compare the forecasts with realizations
Today’s Schedule

- What is Forecasting?
- Point Forecasting
- Linear Forecasting Models
- Forecast Selection: BIC, AIC, $AIC^c$, Mallows, Robust Mallows, FPE, Cross-Validation, PLS, LASSO
- Leading Indicators
- Variance Forecasting
- Interval Forecasting
Example 1

- U.S. Quarterly Real GDP
  - 1960:1-2012:1
Figure: U.S. Real Quarterly GDP
Transformations

- It is mathematically equivalent to forecast $y_{n+h}$ or any monotonic transformation of $y_{n+h}$ and lagged values.
  - It is equivalent to forecast the level of GDP, its logarithm, or percentage growth rate
  - Given a forecast of one, we can construct the forecast of the other.

- Statistically, it is best to forecast a transformation which is close to iid
  - For output and prices, this typically means forecasting growth rates
  - For rates, typically means forecasting changes
Annualized Growth Rate

\[ y_t = 400(\log(Y_t) - \log(Y_{t-1})) \]
Figure: U.S. Real GDP Quarterly Growth
Example 2

- U.S. Monthly 10-Year Treasury Bill Rate
  - 1960:1-2012:4
Figure: U.S. 10-Year Treasury Rate
Monthly Change

$$y_t = Y_t - Y_{t-1}$$
Figure: U.S. 10-Year Treasury Rate Change
Notation

- $y_t$: time series to forecast
- $n$: last observation
- $n + h$: time period to forecast
- $h$: forecast horizon
  - We often want to forecast at long, and multiple, horizons
  - For the first days we focus on one-step ($h = 1$) forecasts, as they are the simplest
- $I_n$: Information available at time $n$ to forecast $y_{n+h}$
  - Univariate: $I_n = (y_n, y_{n-1}, ...)$
  - Multivariate: $I_n = (x_n, x_{n-1}, ...)$ where $x_t$ includes $y_t$, “leading indicators”, covariates, dummy indicators
When we say we want to forecast $y_{n+h}$ given $I_n$,

- We mean that $y_{n+h}$ is uncertain.
- $y_{n+h}$ has a (conditional) distribution $y_{n+h} \mid I_n \sim F(y_{n+h} \mid I_n)$

A complete forecast of $y_{n+h}$ is the conditional distribution $F(y_{n+h} \mid I_n)$ or density $f(y_{n+h} \mid I_n)$

$F(y_{n+h} \mid I_n)$ contains all information about the unknown $y_{n+h}$

Since $F(y_{n+h} \mid I_n)$ is complicated (a distribution) we typically report low dimensional summaries, and these are typically called forecasts.
Standard Forecast Objects

- Point Forecast
- Variance Forecast
- Interval Forecast
- Density forecast
- Fan Chart

All of these forecast objects are features of the conditional distribution.

Today, we focus on point forecasts.
Point Forecasts

- $f_{n+h|h}$, the most common forecast object
- “Best guess” for $y_{n+h}$ given the distribution $F(y_{n+h}|I_n)$
- We can measure its accuracy by a loss function, typically squared error

$$\ell(f, y) = (y - f)^2$$

- The risk is the expected loss

$$E_n \ell(f, y_{n+h}) = E((y_{n+h} - f)^2 | I_n)$$

- The “best” point forecast is the one with the smallest risk

$$f = \arg\min_f E((y_{n+h} - f)^2 | I_n)$$

$$= E(y_{n+h}|I_n)$$

- Thus the optimal point forecast is the true conditional expectation
- Point forecasts are estimates of the conditional expectation
Estimation

- The conditional distribution $F(y_{n+h} | I_n)$ and ideal point forecast $E(y_{n+h} | I_n)$ are unknown.
- They need to be estimated from data and economic models.
- Estimation involves:
  - Approximating $E(y_{n+h} | I_n)$ with a parametric family.
  - Selecting a model within this parametric family.
  - Selecting a sample period (window width).
  - Estimating the parameters.
- The goal of the above steps is not to uncover the “true” $E(y_{n+h} | I_n)$, but to construct a good approximation.
Information Set

- What variables are in the information set $I_n$?
- All past lags
  - $I_n = (x_n, x_{n-1}, ...)$
- What is $x_t$?
  - Own lags, “leading indicators”, covariates, dummy indicators
Markov Approximation

- \( E (y_{n+1} | l_n) = E (y_{n+1} | x_n, x_{n-1}, ...) \)
  - Depends on infinite past

- We typically approximate the dependence on the infinite past with a Markov (finite memory) approximation

- For some \( p \),

  \[
  E (y_{n+1} | x_n, x_{n-1}, ...) \approx E (y_{n+1} | x_n, ..., x_{n-p})
  \]

- This should not be interpreted as true, but rather as an approximation.
Linear Approximation

- While the true $E (y_{n+1} \mid x_n, \ldots, x_{n-p})$ is probably a nasty non-linear function, we typically approximate it by a linear function
  
  $$E (y_{n+1} \mid x_n, \ldots, x_{n-p}) \approx \beta_0 + \beta'_1 x_n + \cdots + \beta'_p x_{n-p}$$
  
  $$= \beta' x_n$$

- Again, this should not be interpreted as true, but rather as an approximation.

- The error is defined as the difference between $y_{n+h}$ and the linear function
  
  $$e_{t+1} = y_{t+1} - \beta' x_t$$
Linear Forecasting Model

- We now have the linear point forecasting model

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

- As this is an approximation, the coefficient and error are defined by projection

\[
\begin{align*}
\beta & = (E (x_t x'_t))^{-1} (E (x_t y_{t+1})) \\
\beta' & = \beta' x_t \\
E (x_t e_{t+1}) & = 0 \\
\sigma^2 & = E (e_{t+1}^2)
\end{align*}
\]

- The conditional variance \( \sigma_t^2 = E (e_{t+1}^2 | I_t) \) may be time-varying
Univariate (Autoregressive) Model

- \( x_t = (y_t, y_{t-1}, \ldots, y_{t-k+1}) \)
- A linear forecasting model is
  \[
  y_{t+1} = \beta_0 + \beta_1 y_t + \beta_2 y_{t-1} + \cdots + \beta_k y_{t-k+1} + e_{t+1}
  \]
- AR(k) – Autoregression of order \( k \)
  - Typical AR(k) models add a stronger assumption about the error \( e_{t+1} \)
    - IID (independent)
    - MDS (unpredictable)
    - White noise (linearly unpredicatable/uncorrelated)
  - These assumptions are convenient for analytic purpose (calculations, simulations)
  - But they are unlikely to be true
    - Making an assumption does not make the assumption true
    - Do not confuse assumptions with truth
Least Squares Estimation

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

\[ \hat{y}_{n+1|n} = \hat{f}_{n+1|n} = \hat{\beta}' x_n \]
GDP Example

- \( y_t = \Delta \log(GDP_t) \), quarterly
- AR(4) (reasonable benchmark for quarterly data)

\[
y_{t+1} = \beta_0 + \beta_1 y_t + \beta_2 y_{t-1} + \beta_3 y_{t-2} + \beta_4 y_{t-3} + e_{t+1}
\]

<table>
<thead>
<tr>
<th></th>
<th>( \hat{\beta} )</th>
<th>( s(\hat{\beta}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1.54</td>
<td>(0.45)</td>
</tr>
<tr>
<td>( \Delta \log(GDP_t) )</td>
<td>0.29</td>
<td>(0.09)</td>
</tr>
<tr>
<td>( \Delta \log(GDP_{t-1}) )</td>
<td>0.18</td>
<td>(0.10)</td>
</tr>
<tr>
<td>( \Delta \log(GDP_{t-2}) )</td>
<td>-0.05</td>
<td>(0.08)</td>
</tr>
<tr>
<td>( \Delta \log(GDP_{t-3}) )</td>
<td>0.06</td>
<td>(0.10)</td>
</tr>
</tbody>
</table>
Point Forecast - GDP Growth

- AR(4)

<table>
<thead>
<tr>
<th>Year</th>
<th>Data</th>
<th>Forecast</th>
<th>Actual</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011:1</td>
<td>0.36</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2011:2</td>
<td>1.33</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2011:3</td>
<td>1.80</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2011:4</td>
<td>2.91</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2012:1</td>
<td>1.84</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2012:2</td>
<td>2.59</td>
<td>1.20</td>
<td></td>
</tr>
</tbody>
</table>
**Interest Rate Example**

- $y_t = \Delta Rate_t$
- AR(12) (reasonable benchmark for monthly data)

<table>
<thead>
<tr>
<th></th>
<th>$\hat{\beta}$</th>
<th>$s(\hat{\beta})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-0.002</td>
<td>0.01</td>
</tr>
<tr>
<td>$\Delta Rate_t$</td>
<td>0.40</td>
<td>0.06</td>
</tr>
<tr>
<td>$\Delta Rate_{t-1}$</td>
<td>-0.26</td>
<td>0.07</td>
</tr>
<tr>
<td>$\Delta Rate_{t-2}$</td>
<td>0.11</td>
<td>0.06</td>
</tr>
<tr>
<td>$\Delta Rate_{t-3}$</td>
<td>-0.07</td>
<td>0.07</td>
</tr>
<tr>
<td>$\Delta Rate_{t-4}$</td>
<td>0.10</td>
<td>0.07</td>
</tr>
<tr>
<td>$\Delta Rate_{t-5}$</td>
<td>-0.08</td>
<td>0.07</td>
</tr>
<tr>
<td>$\Delta Rate_{t-6}$</td>
<td>-0.05</td>
<td>0.06</td>
</tr>
<tr>
<td>$\Delta Rate_{t-7}$</td>
<td>-0.09</td>
<td>0.06</td>
</tr>
<tr>
<td>$\Delta Rate_{t-8}$</td>
<td>-0.01</td>
<td>0.07</td>
</tr>
<tr>
<td>$\Delta Rate_{t-9}$</td>
<td>0.03</td>
<td>0.07</td>
</tr>
<tr>
<td>$\Delta Rate_{t-10}$</td>
<td>0.09</td>
<td>0.07</td>
</tr>
<tr>
<td>$\Delta Rate_{t-11}$</td>
<td>-0.08</td>
<td>0.06</td>
</tr>
</tbody>
</table>
## Point Forecast - 10-year Treasury Rate

- **AR(12)**

<table>
<thead>
<tr>
<th>Year</th>
<th>Data Level</th>
<th>Change</th>
<th>Forecast Level</th>
<th>Change</th>
<th>Actual Level</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012:1</td>
<td>1.97</td>
<td>-0.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2012:2</td>
<td>1.97</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2012:3</td>
<td>2.17</td>
<td>0.20</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2012:4</td>
<td>2.05</td>
<td>-0.12</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2012:5</td>
<td>1.93</td>
<td>-0.12</td>
<td>1.82</td>
<td>-0.23</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Forecast Selection
Forecast Selection

- We used (arbitrarily) an AR(4) for GDP, and an AR(12) for the 10-year rate.
- The forecasts will be sensitive to this choice.
- GDP Example

<table>
<thead>
<tr>
<th>Model</th>
<th>Forecast</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR(0)</td>
<td>2.99</td>
</tr>
<tr>
<td>AR(1)</td>
<td>2.59</td>
</tr>
<tr>
<td>AR(2)</td>
<td>2.65</td>
</tr>
<tr>
<td>AR(3)</td>
<td>2.68</td>
</tr>
<tr>
<td>AR(4)</td>
<td>2.59</td>
</tr>
<tr>
<td>AR(5)</td>
<td>2.83</td>
</tr>
<tr>
<td>AR(6)</td>
<td>2.83</td>
</tr>
<tr>
<td>AR(7)</td>
<td>2.83</td>
</tr>
<tr>
<td>AR(8)</td>
<td>2.78</td>
</tr>
<tr>
<td>AR(9)</td>
<td>2.87</td>
</tr>
<tr>
<td>AR(10)</td>
<td>2.87</td>
</tr>
<tr>
<td>AR(11)</td>
<td>2.91</td>
</tr>
<tr>
<td>AR(12)</td>
<td>3.45</td>
</tr>
</tbody>
</table>
Forecast Selection - Big Picture

- What is the goal?
  - Accurate Forecasts
    - Low Risk (low MSFE)
- Finding the “true” model is irrelevant
  - The true model may be an AR(∞) or have a very large number of non-zero coefficients
Testing

- It is common to use statistical tests to select empirical models
- This is inappropriate
  - Tests answer the scientific question: Is there sufficient evidence to reject the hypothesis that this coefficient is zero?
  - Tests are not designed to answer the question: Which estimate yields the better forecast?
- This is not a minor issue
  - Lengthy statistics literature documenting the poor properties of "post selection" estimators.
  - Estimators based on testing have particularly bad properties
- Tests are appropriate for answering scientific questions about parameters
- Standard errors are appropriate for measuring estimation precision
- For model selection, we want something different
Model Selection: Framework

- Set of estimates (models)
  - $\hat{\beta}(m)$, $m = 1, \ldots, M$
- Corresponding forecasts $\hat{f}_{n+1|n}(m)$
- There is some population criterion $C(m)$ which evaluates the accuracy of $\hat{f}_{n+1|n}(m)$
  - $m_0 = \arg\min_m C(m)$ is infeasible best estimator
- There is a sample estimate $\hat{C}(m)$ of $C(m)$
- $\hat{m} = \arg\min_m \hat{C}(m)$ is empirical analog of $m_0$
- $\hat{\beta}(\hat{m})$ is selected estimator
- $\hat{f}_{n+1|n}(\hat{m})$ selected forecast
Point Forecast and MSFE

- Given an estimate $\hat{\beta}(m)$ of $\beta$, the point forecast for $y_{n+1}$ is
  \[ f_{n+1|n} = \hat{\beta}(m)'x_n \]

- The forecast error is
  \[ y_{n+1} - f_{n+1|n} = x_n'\beta + e_{t+1} - x_n'\hat{\beta}(m) \]
  \[ = e_{n+1} - x_n'\left(\hat{\beta}(m) - \beta\right) \]

- The mean-squared-forecast-error (MSFE) is
  \[ MSFE(m) = E\left(e_{n+1} - x_n'\left(\hat{\beta}(m) - \beta\right)\right)^2 \]
  \[ \approx \sigma^2 + E\left((\hat{\beta}(m) - \beta)' Q(m) \left(\hat{\beta}(m) - \beta\right)\right) \]
  where $Q(m) = E(x_nx_n')$.

- A good forecast has low MSFE.
Selection Criterion

- Bayesian Information Criterion (BIC)
  - \( C(m) = P(m \text{ is true}) \)

- Akaike Information Criterion (AIC), Corrected AIC (AIC\(_c\))
  - \( C(m) = KLIC \)

- Mallows, Predictive Least Squares, Final Prediction Error, Leave-one-out Cross Validation:
  - \( C(m) = MSFE \)

- LASSO
  - Penalized LS
Important: Sample must be constant when comparing models

- This requires careful treatment of samples
- Suppose you observe $y_t, t = 1, ..., n$
- Estimation of an AR($k$) requires $k$ initial conditions, so the effective sample is for observations $t = 1 + k, ..., n$
- The sample varies with $k$, sample size is $n - k$
- For valid comparison of AR($k$) models for $k = 1, ..., K$
  - Fix sample with observations $t = 1 + K, ..., n$
  - $n - K$ observations
  - Estimate all AR($k$) models using this same $n - K$ observations
Bayesian Information Criterion (BIC)

- $M$ models, equal prior probability that each is the “true” model
- Compute posterior probability that model $m$ is true, given data
- Schwarz showed that in the normal linear regression model the posterior probability is proportional to

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

$$BIC(m) = n \log \hat{\sigma}^2(m) + \log(n) k(m)$$

where

- $k(m) =$ # of parameters
- $\hat{\sigma}^2(m) = n^{-1} \sum_{t=0}^{n-1} \hat{e}^2_{t+1}(m) =$ MLE estimate of $\sigma^2$ in model $m$

- The model with highest probability maximizes $p(m)$, or equivalently minimizes $BIC(m)$
Bayesian Information Criterion - Properties

- **Consistent**
  - If true model is finite dimensional, BIC will identify it asymptotically

- **Conservative**
  - Tends to pick small models

- **Inefficient in nonparametric settings**
  - If there is no true finite-dimensional model, BIC is sub-optimal
  - It does not select a finite-sample optimal model

- We are not interested in “truth”, rather we want good performance
Akaike Information Criterion (AIC)

- Estimates Kullback-Leibler information criterion (KLIC) distance between true and estimated density
In the linear regression model

\[ AIC = 2L(\hat{\theta}) + 2k \]
\[ = n \log \hat{\sigma}^2(m) + 2k(m) \]

Similar in form to BIC, but “2” replaces \( \log(n) \)

Picking a model with the smallest AIC is picking the model with the smallest estimated KLIC.
Corrected AIC

In the normal linear regression model, Hurvich-Tsai (1989) calculated the exact AIC

\[ AIC_c(m) = AIC(m) + \frac{2k(m)(k(m) + 1)}{n - k(m) - 1} \]

- Works better in finite samples than uncorrected AIC
- It is an exact correction when the true model is a linear regression, not time series, with iid normal errors.
- In time-series or non-normal errors, it is not an exact correction.
Comments on AIC Selection

- Widely used, partially because of its simplicity
- Full justification requires correct specification
  - normal linear regression
- Critical specification assumption: homoskedasticity
  - AIC is a biased estimate of KLIC under heteroskedasticity
- Criterion: KLIC
  - Not a natural measure of forecast accuracy.
Mallows Criterion

\[ C_n(m) = \tilde{\sigma}^2(m) + \frac{2}{n} \tilde{\sigma}^2 k(m) \]

- Uses a preliminary estimate \( \tilde{\sigma}^2 \) of the variance
- \( C_n(m) \) is an (approximately) unbiased estimate of the MSFE under homoskedasticity and one-step forecasting
- Model \( m \) which minimizes Mallows criterion is an estimate of the lowest MSFE model
Final Prediction Error (FPE) Criterion

\[ FPE_n(m) = \hat{\sigma}^2(m) \left( 1 + \frac{2}{n} k(m) \right) \]
Relation between Mallows, FPE, and Akaike

- Take log of FPE and multiply by $n$

$$n \log \left( FPE_n(m) \right) = n \log \left( \hat{\sigma}^2(m) \right) + n \log \left( 1 + \frac{2}{n} k(m) \right)$$

$$\approx n \log \left( \hat{\sigma}^2(m) \right) + 2k(m)$$

$$= AIC(m)$$

- Thus Mallows, FPE and Akaike model selection are quite similar

- Mallows, FPE, and $\exp \left( AIC(m) / n \right)$ are estimates of MSFE under homoskedasticity
Robust Mallows

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

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

\[ \hat{\Omega}(m) = \frac{1}{n} \sum_{t=0}^{n-1} x_t x'_t \tilde{e}_{t+1}^2 \]

where \( \tilde{e}_{t+1} \) is residual from a preliminary estimate

- An estimate of MSFE, robust to heteroskedasticity and serial correlation (similar to HAC standard errors)
Cross-Validation

- Leave-one-out estimator and prediction residual

\[ \hat{\beta}_{-t}(m) = \left( \sum_{j \neq t} x_j(m)x_j(m)' \right)^{-1} \left( \sum_{j \neq t} x_j(m)y_{j+1} \right) \]

\[ \tilde{e}_{t+1}(m) = y_{t+1} - \hat{\beta}_{-t}(m)'x_t(m) \]

- \( \tilde{e}_{t+1}(m) \) is a forecast error based on estimation without observation \( t \)

- \( E \left( \tilde{e}_{t+1}(m)^2 \right) \approx MSFE_n(m) \)

- \( CV_n(m) = \frac{1}{n} \sum_{t=0}^{n-1} \tilde{e}_{t+1}(m)^2 \) is an estimate of \( MSFE_n(m) \)

- Called the leave-one-out cross-validation (CV) criterion

- Similar to Robust Mallows criterion
Comments on CV Selection

- Selecting one-step forecast models by cross-validation is computationally simple, generally valid, and robust to heteroskedasticity.
- Does not require correct specification.
- Similar to robust Mallows.
- Similar to Mallows, AIC and FPE under homoskedasticity.
- Conceptually easy to generalize beyond least-squares estimation.
Predictive Least Squares (Out-of-Sample MSFE)

- Sequential estimates

\[
\hat{\beta}_t(m) = \left( \sum_{j=0}^{t-1} x_j(m)x_j(m)' \right)^{-1} \left( \sum_{j=0}^{t-1} x_j(m)y_{j+1} \right)
\]

- Sequential prediction residuals

\[
\bar{e}_{t+1}(m) = y_{t+1} - \hat{\beta}_t(m)'x_t(m)
\]

- Predictive Least Squares. For some \( P \)

\[
PLS_n(m) = \frac{1}{P} \sum_{t=n-P}^{n-1} \bar{e}_{t+1}(m)^2
\]

- Major Difficulty: PLS very sensitive to \( P \)
Comments on Predictive Least Squares

- Conceptually simple, easy to generalize beyond least-squares
  - Can be applied to actual forecasts, without need to know forecast method
- $\bar{e}_{t+1}(m)$ are fully valid prediction errors
- Possibly more robust to structural change than CV
  - Intuitive, but this claim has not been formally justified
- Very common in applied forecasting
  - Frequently asserted as “empirical performance”
- On the negative side, PLS over-estimates MSFE
  - $\bar{e}_{t+1}(m)$ is a prediction error from a sample of length $t < n$
  - PLS will tend to be overly-parsimonious
  - Very sensitive to number of pseudo out-of-sample observations $P$
Theory of Optimal Selection

- $MSFE_n(m)$ is the MSFE from model $m$
- $\inf_m MSFE_n(m)$ is the (infeasible) best MSFE
- Let $\hat{m}$ be the selected model
- Let $MSFE_n(\hat{m})$ denote the MSFE using the selected estimator
- We say that selection is asymptotically optimal if

$$\frac{MSFE_n(\hat{m})}{\inf_m MSFE_n(m)} \xrightarrow{p} 1$$
Theory of Optimal Selection

- A series of papers have shown that AIC, Mallows, FPE are asymptotically optimal for selection

- Assumptions
  - Autoregressions
  - Errors are iid, homoskedastic
  - True model is AR(∞)

- Shibata (Annals, 1980), Ching-Kang Ing with co-authors (2003, 2005, etc)

- Proof Method: Show that the selection criterion is uniformly close to MSFE
Theory of Optimal Selection - Regression Case

- In regression (iid data) case
- AIC, Mallows, FPE, CV are asymptotically optimal for selection under homoskedasticity
- CV is asymptotically optimal for selection under heteroskedasticity
Forecast Selection - Summary

- Testing inappropriate for forecast selection
- Feasible selection criteria: BIC, AIC, AIC$_c$, Mallows, Robust Mallows, FPE, PLS, CV
- Valid comparisons require holding sample constant across models
- All methods except CV and PLS require conditional homoskedasticity
- PLS sensitive to choice of $P$
- BIC appropriate when true structure is sparse
- CV quite general and flexible
  - Recommended method
GDP Example

Methods: BIC, $AIC_c$, Robust Mallows, CV

<table>
<thead>
<tr>
<th>Model</th>
<th>BIC</th>
<th>$AIC_c$</th>
<th>$C_n^*$</th>
<th>CV</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR(1)</td>
<td>473</td>
<td>466</td>
<td>10.7</td>
<td>10.7</td>
</tr>
<tr>
<td>AR(2)</td>
<td>472</td>
<td>462</td>
<td>10.6</td>
<td>10.5</td>
</tr>
<tr>
<td>AR(3)</td>
<td>477</td>
<td>464</td>
<td>10.7</td>
<td>10.7</td>
</tr>
<tr>
<td>AR(4)</td>
<td>481</td>
<td>465</td>
<td>10.8</td>
<td>10.8</td>
</tr>
<tr>
<td>AR(5)</td>
<td>483</td>
<td>464</td>
<td>10.8</td>
<td>10.8</td>
</tr>
<tr>
<td>AR(6)</td>
<td>489</td>
<td>466</td>
<td>11.0</td>
<td>10.9</td>
</tr>
<tr>
<td>AR(7)</td>
<td>494</td>
<td>468</td>
<td>11.1</td>
<td>11.1</td>
</tr>
<tr>
<td>AR(8)</td>
<td>498</td>
<td>470</td>
<td>11.3</td>
<td>11.2</td>
</tr>
<tr>
<td>AR(9)</td>
<td>500</td>
<td>469</td>
<td>11.3</td>
<td>11.2</td>
</tr>
<tr>
<td>AR(10)</td>
<td>505</td>
<td>471</td>
<td>11.4</td>
<td>11.4</td>
</tr>
<tr>
<td>AR(11)</td>
<td>511</td>
<td>473</td>
<td>11.5</td>
<td>11.5</td>
</tr>
<tr>
<td>AR(12)</td>
<td>511</td>
<td>471</td>
<td>11.4</td>
<td>11.3</td>
</tr>
</tbody>
</table>

Methods select AR(2)
## 10-Year Treasury Rate

<table>
<thead>
<tr>
<th>Model</th>
<th>BIC</th>
<th>(\text{AIC}_c)</th>
<th>(C_n^*)</th>
<th>CV</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR(1)</td>
<td>-1518</td>
<td>-1527</td>
<td>0.0798</td>
<td>0.0798</td>
</tr>
<tr>
<td>AR(2)</td>
<td>-1541*</td>
<td>-1554</td>
<td>0.0768*</td>
<td>0.0768*</td>
</tr>
<tr>
<td>AR(3)</td>
<td>-1538</td>
<td>-1555</td>
<td>0.0769</td>
<td>0.0769</td>
</tr>
<tr>
<td>AR(4)</td>
<td>-1532</td>
<td>-1554</td>
<td>0.0773</td>
<td>0.0773</td>
</tr>
<tr>
<td>AR(6)</td>
<td>-1531</td>
<td>-1561</td>
<td>0.0772</td>
<td>0.0770</td>
</tr>
<tr>
<td>AR(8)</td>
<td>-1522</td>
<td>-1562</td>
<td>0.0777</td>
<td>0.0774</td>
</tr>
<tr>
<td>AR(10)</td>
<td>-1513</td>
<td>-1561</td>
<td>0.0784</td>
<td>0.0781</td>
</tr>
<tr>
<td>AR(12)</td>
<td>-1506</td>
<td>-1563</td>
<td>0.0790</td>
<td>0.0787</td>
</tr>
<tr>
<td>AR(20)</td>
<td>-1471</td>
<td>-1561</td>
<td>0.0810</td>
<td>0.0800</td>
</tr>
<tr>
<td>AR(22)</td>
<td>-1470</td>
<td>-1570*</td>
<td>0.0810</td>
<td>0.0800</td>
</tr>
<tr>
<td>AR(24)</td>
<td>-1458</td>
<td>-1565</td>
<td>0.0810</td>
<td>0.0810</td>
</tr>
</tbody>
</table>

Mallows, \(\text{AIC}_c\), FPE select AR(22)
Robust Mallows, CV select AR(2)
Difference due to conditional heteroskedasticity
AR(2) through AR(6) near equivalent with respect to \(C_n^*\) and CV
Point Forecast - GDP Growth

- AR(2)

<table>
<thead>
<tr>
<th>Year</th>
<th>Data</th>
<th>Forecast</th>
<th>Actual</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011:1</td>
<td>0.36</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2011:2</td>
<td>1.33</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2011:3</td>
<td>1.80</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2011:4</td>
<td>2.91</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2012:1</td>
<td>1.84</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2012:2</td>
<td>2.65</td>
<td>1.20</td>
<td></td>
</tr>
</tbody>
</table>
### Point Forecast - 10-year Treasury Rate

- **AR(2)**

<table>
<thead>
<tr>
<th>Year</th>
<th>Data Level</th>
<th>Change</th>
<th>Forecast Level</th>
<th>Change</th>
<th>Actual Level</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012:1</td>
<td>1.97</td>
<td>-0.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2012:2</td>
<td>1.97</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2012:3</td>
<td>2.17</td>
<td>0.20</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2012:4</td>
<td>2.05</td>
<td>-0.12</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2012:5</td>
<td></td>
<td>1.96</td>
<td>-0.09</td>
<td></td>
<td>1.82</td>
<td>-0.23</td>
</tr>
</tbody>
</table>
Recall, the ideal forecast is

$$E(y_{n+1}|I_n) = E(y_{n+1}|x_n, x_{n-1}, \ldots)$$

where $I_n$ contains all information

$x_n = \text{lags} + \text{leading indicators}$

- Variables which help predict $y_{t+1}$
- We have focused on univariate lags
- Typically more information in related series
- Which?
Good Leading Indicators

- Measured quickly
- Anticipatory
- Varies by forecast variable
Interest Rate Spreads

- Difference between Long and Short Rate
- Measured immediately
- Indicate monetary policy, aggregate demand
- Term Structure of Interest Rates:
  - Long Rate is the market expectation of the average future short rates
  - Spread is the market expectation of future short rates
- I use U.S. Treasury rates, difference between 10-year and 3-month
Figure: 10-Year and 3-Month T-Bill Rates
Figure: Term Spread
High Yield Spread

- “Riskless” rate: U.S. Treasury
- Low-risk rate: AAA grade corporate bond
- High Yield rate: Low grade corporate bond
- Theory: high-yield rate includes premium for probability of default
- Low grade bond rates increase with probability of default – when real activity is expected to fall
- Spread: Difference between corporate bond rates
- I use difference between AAA and BAA bond rates
Figure: AAA and BAA Corporate Bond Rates
Figure: High Yield Spread
Construction Indicators

- Building Permits
- Housing Starts
- Anticipate construction spending
Figure: Housing Starts, Building Permits
Mixed Frequency Data

- U.S. GDP is measured quarterly
- Interest rates: Daily
- Permits: Monthly
- Simplest approach: Quarterly aggregation
  - Aggregate (average) daily and monthly variables to quarterly level
- Mixed Frequency approach
  - Use lower frequency data as predictors
- For now, we use aggregate (quarterly) data
Timing

- Variables reported in separate sequences
- Should we use only "quarter 1" variables to forecast "quarter 2"?
- Or should we use whatever is available?
  - E.g., use quarter 2 interest rates to forecast quarter 1 GDP?
- Let’s use quarter 1 data to forecast quarter 2
## Models Selection by CV

- All estimates include intercept plus two lags of GDP growth

<table>
<thead>
<tr>
<th>Model</th>
<th>CV</th>
<th>Forecast</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spread</td>
<td>10.4</td>
<td>2.8</td>
</tr>
<tr>
<td>HY Spread</td>
<td>10.6</td>
<td>2.5</td>
</tr>
<tr>
<td>Housing Starts</td>
<td>10.3</td>
<td>1.4</td>
</tr>
<tr>
<td>Building Permits</td>
<td>10.3</td>
<td>1.7</td>
</tr>
<tr>
<td>Sp+HY</td>
<td>10.3</td>
<td>2.7</td>
</tr>
<tr>
<td>Sp+HS</td>
<td>10.02</td>
<td>1.5</td>
</tr>
<tr>
<td>Sp+BP</td>
<td>10.1</td>
<td>1.9</td>
</tr>
<tr>
<td>HY+HS</td>
<td>10.4</td>
<td>1.4</td>
</tr>
<tr>
<td>HY+BP</td>
<td>10.4</td>
<td>1.6</td>
</tr>
<tr>
<td>HS+BP</td>
<td>10.4</td>
<td>1.4</td>
</tr>
<tr>
<td><strong>Sp+HY+HS</strong></td>
<td><strong>10.00</strong></td>
<td><strong>1.3</strong></td>
</tr>
<tr>
<td>Sp+HY+BP</td>
<td>10.1</td>
<td>1.7</td>
</tr>
<tr>
<td>Sp+HS+BP</td>
<td>10.05</td>
<td>1.3</td>
</tr>
<tr>
<td>HY+HS+BP</td>
<td>10.5</td>
<td>1.3</td>
</tr>
<tr>
<td>Sp+HY+HS+BP</td>
<td>10.02</td>
<td>1.1</td>
</tr>
</tbody>
</table>
CV-Selected Forecast: 1.3%
Actual: 1.2%
Coefficient Estimates

<table>
<thead>
<tr>
<th>Term</th>
<th>$\hat{\beta}$</th>
<th>$s(\hat{\beta})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta \log(GDP_{t+1})$</td>
<td>$-0.33$</td>
<td>$(1.03)$</td>
</tr>
<tr>
<td>Intercept</td>
<td>$-0.33$</td>
<td>$(1.03)$</td>
</tr>
<tr>
<td>$\Delta \log(GDP_t)$</td>
<td>$0.16$</td>
<td>$(0.10)$</td>
</tr>
<tr>
<td>$\Delta \log(GDP_{t-1})$</td>
<td>$0.09$</td>
<td>$(0.10)$</td>
</tr>
<tr>
<td>Bond Spread$_t$</td>
<td>$0.61$</td>
<td>$(0.23)$</td>
</tr>
<tr>
<td>High Yield Spread$_t$</td>
<td>$-1.10$</td>
<td>$(0.75)$</td>
</tr>
<tr>
<td>Housing Starts$_t$</td>
<td>$1.86$</td>
<td>$(0.65)$</td>
</tr>
</tbody>
</table>
Variance Forecasting
Variance Forecasts

- Forecast uncertainty
  - Point forecasts insufficient!

\[ \sigma^2_{t+1} = \text{var} \left( y_{t+1} | l_t \right) \]

- In the model \( y_{t+1} = \beta' x_t + e_{t+1} \)
  - \( \sigma^2_{t+1} = \text{var} \left( e_{t+1} | l_n \right) = \mathbb{E} \left( e^2_{t+1} | l_t \right) \)
10-Year Bond Rate

- Prediction Residuals
- Squares
Figure: Leave-One-Out Prediction Residuals
Figure: Squared Prediction Residuals
Variance Forecast Methods

- **Constant Variance** $\sigma^2_t = \sigma^2$
  - Uncertainty not state-dependent

- **GARCH**
  - Common in financial data
  - Estimated by MLE

- **Regression Approach**
  - $\sigma^2_t = E(e^2_{t+1} | l_n) \approx \alpha' x_t$
2-Step Variance Estimation

- Start with residuals $\hat{e}_{t+1}$
  - Better choice: leave-one-out residuals $\tilde{e}_{t+1}$
- Estimate variance model (constant, ARCH, or regression)
- Obtain $\hat{\sigma}_n^2$ from fitted model
Which Residuals?

- Least-squares residual variance biased toward zero
  - Forecast variance biased towards zero
- Leave-one-out residual variance estimates out-of-sample MSFE
  - Better choice

\[
\tilde{e}_{t+1}(m) = y_{t+1} - \hat{\beta}_{-t}(m)'x_t(m)
\]

\[
\hat{\beta}_{-t}(m) = \left( \sum_{j \neq t} x_j(m)x_j(m)' \right)^{-1} \left( \sum_{j \neq t} x_j(m)y_{j+1} \right)
\]
Constant Variance Model

- $\sigma_t^2 = \sigma^2$
- $\hat{\sigma}_n^2 = \hat{\sigma}^2 = \frac{1}{n-1} \sum_{t=1}^{n-1} \hat{e}_{t+1}^2$
Regression Variance Model

- $\sigma_t^2 \approx \alpha'x_t$
- $e_{t+1}^2 = \alpha'x_t + \eta_t$
- $\hat{\alpha} = (\sum_{t=1}^{n-1} x_t x'_t)^{-1} (\sum_{t=1}^{n-1} x_t \tilde{e}_{t+1}^2)$
- $\hat{\sigma}_n^2 = \hat{\alpha}'x_n$
  - Easy, but not constrained to $(0, \infty)$
GARCH Models

- \( \sigma_t^2 = \omega + \beta \sigma_{t-1}^2 + \alpha e_t^2 \)
- Conditional variance of \( e_{t+1} \)
- Specifies conditional variance as function of recent squared innovations
- Large innovations (in magnitude) raise conditional variance
- Lagged variance smooths \( \sigma_t^2 \)
- Non-negativity constraints: \( \omega > 0, \beta \geq 0, \alpha > 0 \)
GARCH with Regressors

\[ \sigma_t^2 = \omega + \beta \sigma_{t-1}^2 + \alpha e_t^2 + \gamma x_t \]

- \( x_t > 0 \) useful to constrain regressor to be positive
Estimation by Quasi-Likelihood

- Numerical optimization
Model Selection

- Model with 2 ARCH lags and 2 regressors

\[
\sigma_t^2 = \omega + \beta \sigma_{t-1}^2 + \alpha_1 e_t^2 + \alpha_2 e_{t-1}^2 + \gamma_1 x_{1t} + \gamma_2 x_{2t}
\]

- How many lags? How many regressors?
- Presence of lagged \( \sigma_{t-1}^2 \) complicates issues
  - \( \beta \) not identified when \( \alpha_1 = \alpha_2 = \gamma_1 = \gamma_2 = 0 \)
  - This means conventional tests and information criterion are not correct when the process is close to constant variance
  - We typically ignore this complication

- Since estimation is nonlinear MLE much of model selection & combination literature is not relevant
  - AIC appropriate
  - Unfortunately, not easy to compute with standard packages
AIC for GARCH models

If model $m$ has parameter vector $\theta(m)$ with $k(m)$ elements

- $AIC(m) = 2\mathcal{L}(\hat{\theta}(m)) + 2k(m)$
- Not standard output
Variance Forecast from GARCH model

- \( \sigma^2_{n+1} = \omega + \beta \sigma^2_n + \alpha_1 e^2_n \)
- \( \hat{\sigma}^2_{n+1} = \hat{\omega} + \hat{\beta} \hat{\sigma}^2_n + \hat{\alpha}_1 \hat{e}^2_n \)
- \( \hat{\sigma}^2_{n+1} \) is estimated conditional variance of \( y_{n+1} \)
- Standard deviation \( \sqrt{\hat{\sigma}^2_{n+1}} \)
Example: 10-Year Bond Rate

GARCH(1,1)

\[ \sigma^2_t = \omega + \alpha e^2_t + \beta \sigma^2_{t-1} \]

| \( \omega \) | 0.0001 | 0.0001 |
| \( \alpha \) | 0.200  | 0.041  |
| \( \beta \)  | 0.835  | 0.025  |
Variance Forecast

- **Conditional variance**
  - $\sigma_{n+1}^2 = 0.054$
  - $\hat{\sigma}_{n+1} = 0.23$

- **Unconditional**
  - $\hat{\sigma}^2 = 0.076$
  - $\hat{\sigma} = 0.28$

- The conditional variance at present is similar, but somewhat smaller than the unconditional
Figure: Estimated Variance
Example: GDP Growth
Figure: GDP: Leave-One-Out Prediction Residuals
Figure: GDP: Squared Prediction Residuals
GARCH(1)

\[ \sigma_t^2 = \omega + \alpha e_t^2 + \beta \sigma_{t-1}^2 \]

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>s.e.</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \omega )</td>
<td>0.81</td>
<td>0.46</td>
</tr>
<tr>
<td>( \alpha )</td>
<td>0.21</td>
<td>0.06</td>
</tr>
<tr>
<td>( \beta )</td>
<td>0.72</td>
<td>0.06</td>
</tr>
</tbody>
</table>

- **Conditional variance**
  - \( \hat{\sigma}_{n+1}^2 = 4.1 \)
  - \( \hat{\sigma}_{n+1} = 2.0 \)

- **Unconditional**
  - \( \hat{\sigma}^2 = 9.8 \)
  - \( \hat{\sigma} = 3.1 \)
Figure: GDP: Estimated Variance
Interval Forecasting
Interval Forecasts

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

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

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
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

\[
\begin{align*}
y_{t+1} &= \mu_t + \sigma_t \varepsilon_{t+1} \\
\mu_t &= E(y_{t+1} | l_t) \\
\sigma^2_t &= \text{var}(y_{t+1} | l_t)
\end{align*}
\]

- 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

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

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

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

- Interval forecasts thus can be summarized by $\mu_n$, $\sigma_n^2$, and $q^\epsilon(\alpha)$
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

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

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

- When the estimated variance is a constant, this is numerically identical to the definition with rescaled errors $\tilde{e}_{t+1}$
Example: Interest Rate Forecast

- $n = 603$ observations
- $\tilde{\varepsilon}_{t+1} = \frac{\tilde{e}_{t+1}}{\tilde{\sigma}_t}$ 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 $= [1.69, 2.25]$
- Actual: $1.82$
Example: GDP

- \( n = 207 \) observations
- \( \hat{e}_{t+1} = \frac{\tilde{e}_{t+1}}{\hat{\sigma}_t} \) from GARCH(1,1) model
- 0.10, 0.25, 0.75, 0.90 quantiles
- \(-1.18, -0.63, 0.57, 1.26\)
- Point Forecast = 1.31
- 50\% Forecast interval = \([0.04, 2.4]\)
- 80\% Forecast interval = \([-1.07, 3.8]\)
- Actual: 1.20
Mean-Variance Model Interval Forecasts - Summary

- The key is to break the distribution into the mean $\mu_t$, variance $\sigma_t^2$ 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 $\mu_n$, $\sigma_n^2$ 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.