Statistics 5350/7110
Forecasting

Lecture 24
Forecast Revisions

Professor Robert Stine

Preliminaries

- Questions?
  - No office hours today
- Final exam
  - Similar to Midterm: short answer, multiple choice
  - Friday, December 13, 3-5 pm
  - JMHH F45
- Quick review
  - Periodogram: Decomposing variation in a time series by frequency
  - Periodogram-based diagnostics: Max and cumulative
  - Spectral density: Alternative description of ARMA models

Today's Topics

- Forecast structure
  - Review of ARMA properties
  - Correlation of forecast errors
  - ARMA models suggest general case

- Recognizing when forecasts are well-behaved
  - Martingale
  - Threshold martingale

- More details, examples from sports betting
  - Threshold Martingales and the Evolution of Forecasts, Foster & Stine (2021), ArXiv

Forecast Structure

Forecast Structure

- Representations of a stationary ARMA process
  - Difference equation

$$X_t = \alpha + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \theta_1 w_{t-1} + w_t$$

  - Autoregression (invertible)

$$X_t = \alpha + w_t + \pi_1 X_{t-1} + \pi_2 X_{t-2} + \cdots$$

  - Moving average

$$X_t = \mu + w_t + \psi_1 w_{t-1} + \psi_2 w_{t-2} + \cdots$$

- Exploiting MA form
  - Convenient since weighted sum of uncorrelated variables (white noise)
  - Forecast construction, presuming you know w_t for $t \leq n$

$$\widehat{X}_{n+s|n} = E\left(X_{n+s} \,|\, X_n, X_{n-1}, \ldots\right) = \mu + \psi_s w_n + \psi_{s+1} w_{n-1} + \cdots$$

  - Forecast error

$$X_{n+s} - \widehat{X}_{n+s|n} = w_{n+s} + \psi_1 w_{n+s-1} + \cdots + \psi_{s-1} w_{n+1}$$

Forecast Revisions

- Forecasts change as target gets closer
  - Same weights on lagged white noise, shifted from unknown to known
  - Forecast lead 3 steps out

$$\begin{aligned} \widehat{X}_{n+3|n} &= \mu + \psi_3 w_n + \psi_4 w_{n-1} + \psi_5 w_{n-2} + \cdots \\ X_{n+3} - \widehat{X}_{n+3|n} &= w_{n+3} + \psi_1 w_{n+2} + \psi_2 w_{n+1} \end{aligned}$$

  - Forecast lead 2 steps out

$$\begin{aligned} \widehat{X}_{n+3|n+1} &= \mu + \psi_2 w_{n+1} + \psi_3 w_n + \psi_4 w_{n-1} + \psi_5 w_{n-2} + \cdots \\ X_{n+3} - \widehat{X}_{n+3|n+1} &= w_{n+3} + \psi_1 w_{n+2} \end{aligned}$$

  - Forecast lead 1 step out

$$\begin{aligned} \widehat{X}_{n+3|n+2} &= \mu + \psi_1 w_{n+2} + \psi_2 w_{n+1} + \psi_3 w_n + \psi_4 w_{n-1} + \psi_5 w_{n-2} + \cdots \\ X_{n+3} - \widehat{X}_{n+3|n+1} &= w_{n+3} \end{aligned}$$

- Correlation of forecasts and errors
  - Overlapping sums of white noise

Example: Forecast Revisions

- AR(1) case, known process
  - Moving average weights are $\psi_j = \phi^j$
- Forecasts change as target gets closer
  - Same weights on lagged white noise, shifted from unknown to known
  - Forecast lead 3

$$\widehat{X}_{n+3|n} = \mu + \phi^3 w_n + \phi^4 w_{n-1} + \phi^5 w_{n-2} + \cdots$$
$$X_{n+3} - \widehat{X}_{n+3|n} = w_{n+3} + \phi\, w_{n+2} + \phi^2 w_{n+1}$$

  - Forecast lead 2

$$\widehat{X}_{n+3|n+1} = \mu + \phi^2 w_{n+1} + \phi^3 w_n + \phi^4 w_{n-1} + \phi^5 w_{n-2} + \cdots$$
$$X_{n+3} - \widehat{X}_{n+3|n+1} = w_{n+3} + \phi\, w_{n+2}$$

  - Forecast lead 1

$$\widehat{X}_{n+3|n+2} = \mu + \phi\, w_{n+2} + \phi^2 w_{n+1} + \phi^3 w_n + \phi^4 w_{n-1} + \phi^5 w_{n-2} + \cdots$$
$$X_{n+3} - \widehat{X}_{n+3|n+1} = w_{n+3}$$

Example: Forecast Revisions

- AR(1) case, known process
  - Moving average weights are $\psi_j = \phi^j$
- Forecasts change as target gets closer
  - Reexpress the predictor
  - Forecast lead 3

$$\widehat{X}_{n+3|n} = \mu + \phi^3(w_n + \phi^4 w_{n-1} + \phi^5 w_{n-2} + \cdots) = \phi^3 X_n$$
$$X_{n+3} - \widehat{X}_{n+3|n} = w_{n+3} + \phi\, w_{n+2} + \phi^2 w_{n+1}$$

  - Forecast lead 2

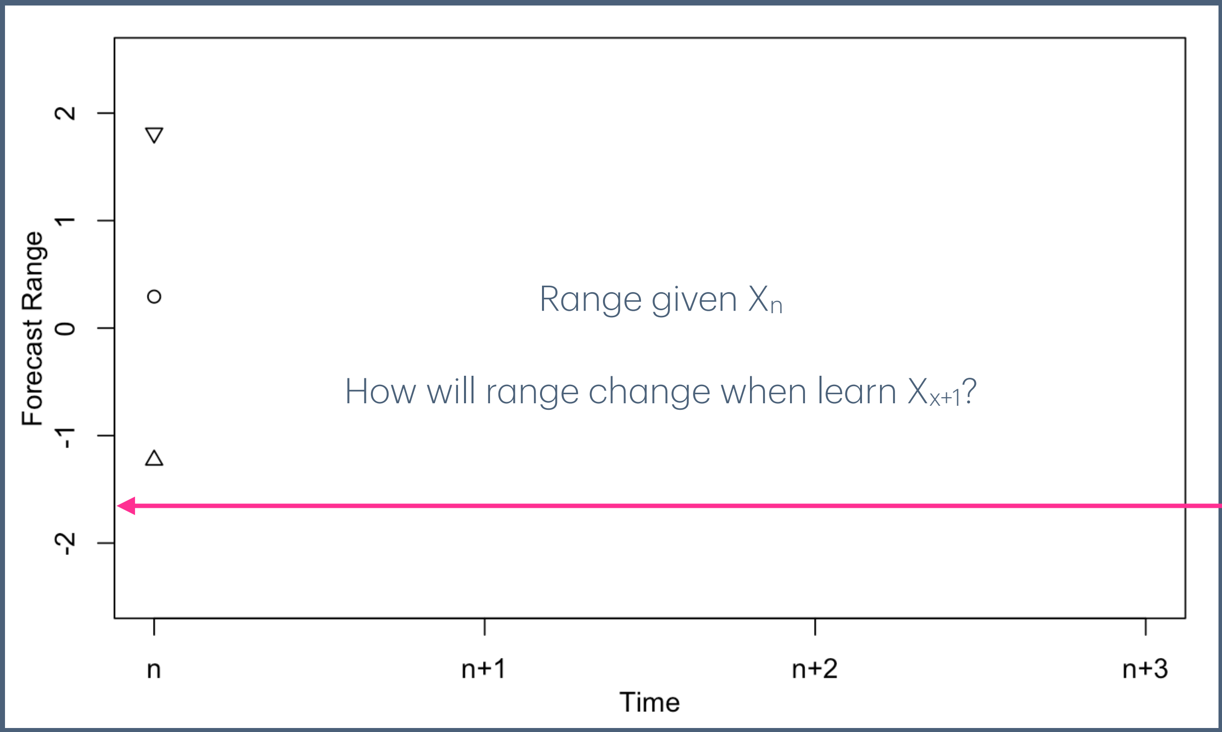
$$\widehat{X}_{n+3|n+1} = \mu + \phi^2 X_{n+1}$$
$$X_{n+3} - \widehat{X}_{n+3|n+1} = w_{n+3} + \phi\, w_{n+2}$$

  - Forecast lead 1

$$\widehat{X}_{n+3|n+2} = \mu + \phi\, X_{n+2}$$
$$X_{n+3} - \widehat{X}_{n+3|n+1} = w_{n+3}$$

Example: Forecast Revisions

- AR(1) case, known process… optimal revisions
  - Moving average weights are $\psi_j = \phi^j$



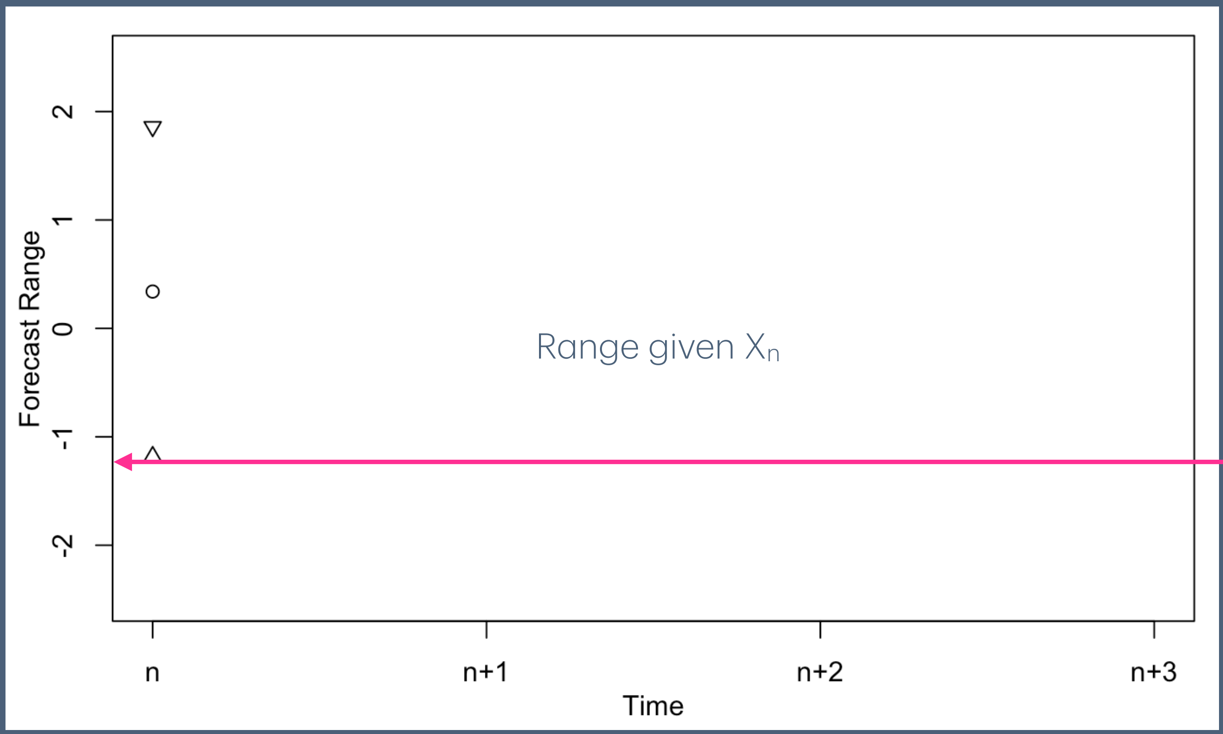
Value of X_{n+4}

Example: Forecast Revisions

- AR(1) case, known process… optimal revisions
  - Moving average weights are $\psi_j = \phi^j$

Same process as prior slide

Independent realization



Value of X_{n+4}

Forecasts and Martingales

Forecast Updates

- Changes to forecast
  - Three steps out

$$X_{n+3} = w_{n+3} + \phi\, w_{n+2} + \phi^2 w_{n+1} + \widehat{X}_{n+3|n}$$

  - Two steps out

$$X_{n+3} = w_{n+3} + \phi_1 w_{n+2} + \widehat{X}_{n+3|n+1}$$

  - One step out

$$X_{n+3} = w_{n+3} + \widehat{X}_{n+3|n+2}$$

- Properties of the changes
  - Variance of changes to location increases as target approaches: information flow
  - Expected values of future changes are zero
  - Variance of remaining error shrinks
- General case
  - Capture these properties without assuming you have a particular model (e.g. AR(1))

Martingale

- Definition
  - Everything known up to time t $\mathscr{F}_t = X_t, X_{t-1}, X_{t-2}, \ldots$
  - Process X_t is a martingale if

$$E\left(X_{t+1} \mid X_t, X_{t-1}, \ldots\right) = E\left(X_{t+1} \mid \mathscr{F}_t\right) = X_t$$

$$E\left(X_{t+s} \mid \mathscr{F}_t\right) = X_t, \quad 0 < s$$

  - Martingale differences are uncorrelated

$$E\left((X_t - X_{t-1})(X_s - X_{s-1})\right) = 0, \quad s < t$$

uncorrelated martingale difference independent

- Main example
  - Accumulating sum: Random walk (without drift)

$$X_t = \sum_{t=1}^{n} w_t \quad \Rightarrow \quad X_{t+1} = X_t + w_t \qquad t = 2, 3, \ldots$$

- Stationary ARMA processes
  - Mean reverting, hence not martingale
  - AR(1) example

$$X_t = \phi X_{t-1} + w_t \quad \Rightarrow \quad E\left(X_{t+1} \mid \mathscr{F}_t\right) = \phi X_t$$

Martingales are special…
You have to make them!

Forecasts as Martingales

- Target
  - Observe time series at times … t, t+1, …, T
  - At various times t < T, we forecast X_T
- Doob's martingale
  - Y is a random variable
  - Define process

$$\widehat{Y}_t = E\left(Y \mid \mathcal{F}_t\right)$$

  This process is a martingale.

Proof is an exercise in conditional expected values

- Forecast martingale
  - Forecasts of X_T as t approaches target time T

$$\widehat{X}_{T|T-s}, \widehat{X}_{T|T-s+1}, \ldots, \widehat{X}_{T|T-1}, \widehat{X}_{T|T} = X_T$$

  - Example of ARMA forecasts for a fixed target

$$\widehat{X}_{T|t} = \psi_{T-t} w_t + \widehat{X}_{T|t-1}$$

  where the increment has mean 0 and are uncorrelated

Threshold Martingale

- Definition
  - Build martingale from probabilities rather than values
  - Pick a target time T and a threshold τ
  - Define martingale

$$p_t = E\left(I(Y_T \le \tau) \,|\, \mathscr{F}_t\right) = P\left(Y_T \le \tau \,|\, \mathscr{F}_t\right)$$

  where $I(\text{statement}) = 1$ if statement is true and 0 otherwise (like a dummy variable)

  - Denote mean $\pi = E\,p_t$
- Properties
  - A martingale by construction
  - Calibrated

$$E\left(p_t \,|\, p_{t-s}\right) = p_{t-s}, \qquad 0 < s$$

  - Variance is known

$$E\left(p_T - \pi\right)^2 = \pi\,(1-\pi)$$

  Leads to notion of "excess volatility" in forecast updates as well as "information gain"

Applying Threshold Martingale

- Definition
  - Pick a target time T and a threshold τ

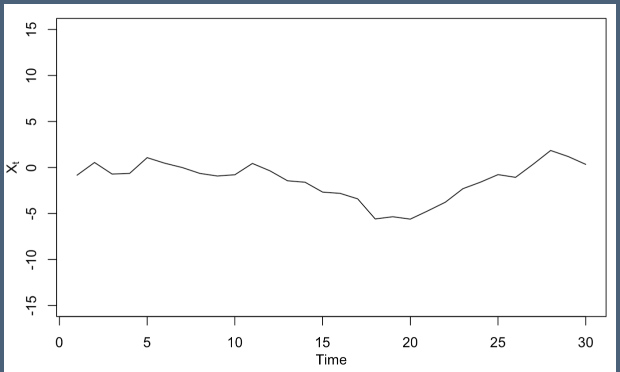
$$p_t = P\left(Y_T \le \tau \,|\, \mathscr{F}_t\right) \qquad \text{with} \qquad \pi = E\, p_t$$

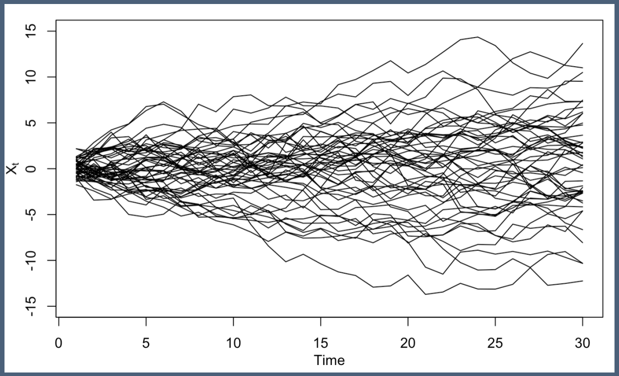
- Properties
  - Calibration $\qquad E\left(p_t \,|\, p_{t-s}\right) = p_{t-s}, \qquad 0 < s$

    Implies that a plot of p_t on prior values should average on diagonal
  - Variation $\qquad E\left(p_T - \pi\right)^2 = \pi\left(1 - \pi\right)$

    Should accumulate toward π(1-π) as forecasts approach target

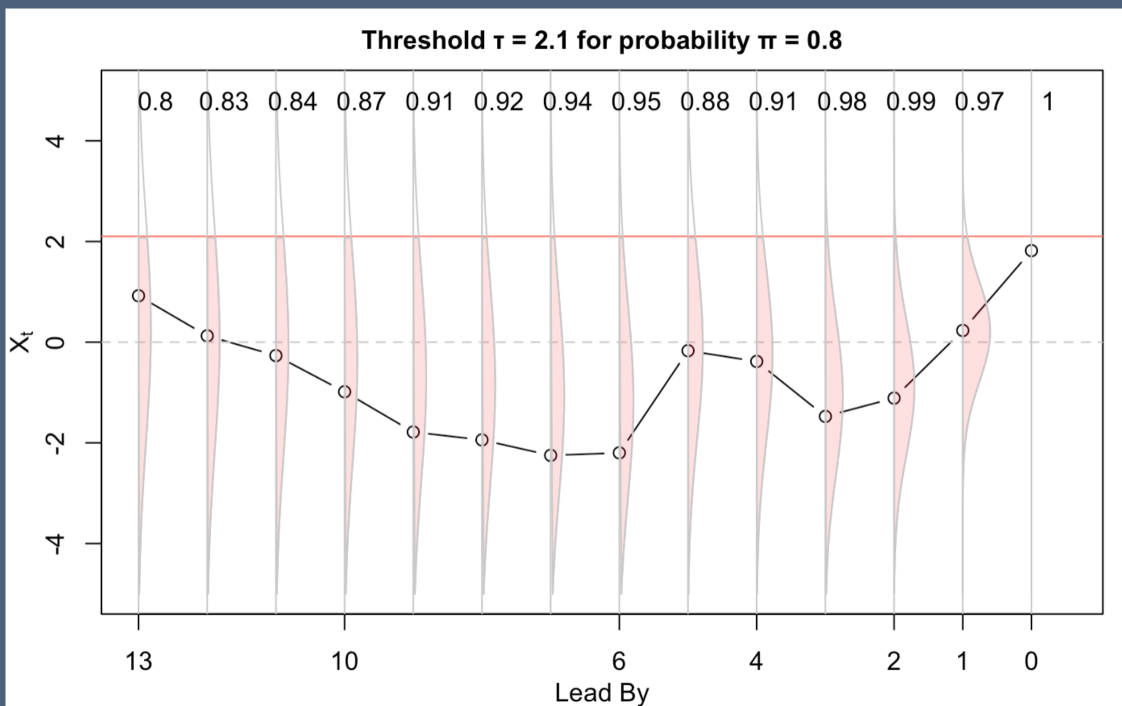
- Concern
  - Hard to recognize RW from seeing one realization
  - Some trend up, some down. Some wander

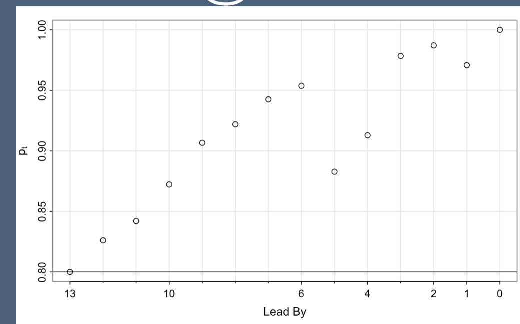


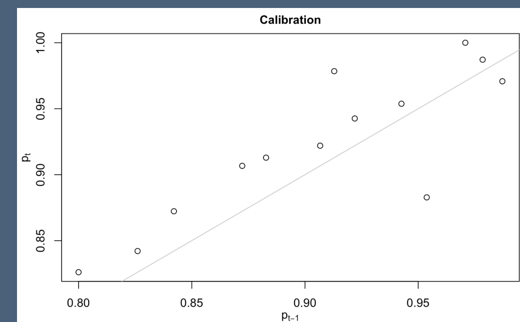


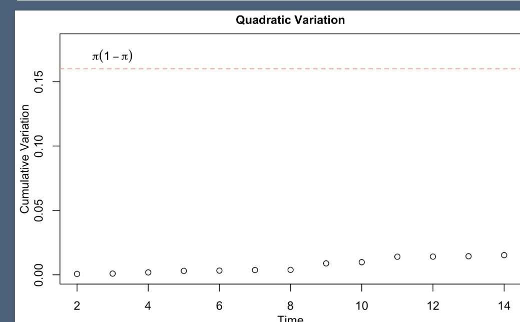
Example of Threshold Martingale

- AR(1) process
  - Points are values of X_n, X_{n+1}, ..., X_T
  - Threshold τ is 80% point of initial forecast distribution at time n+1
  - Shaded regions determine p_t



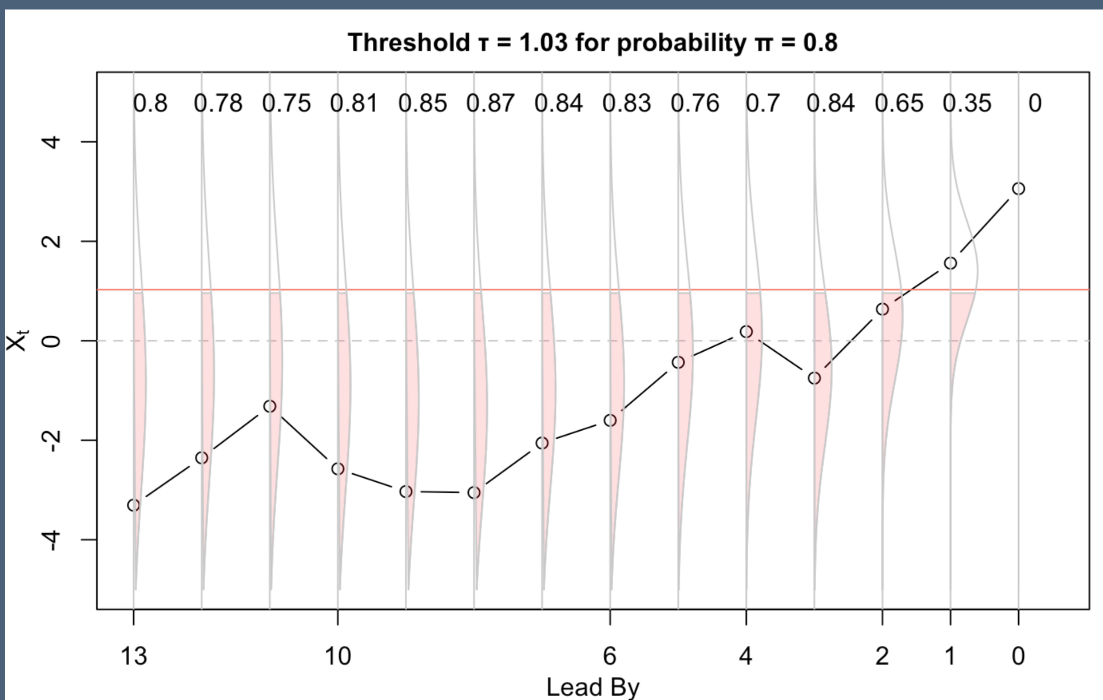


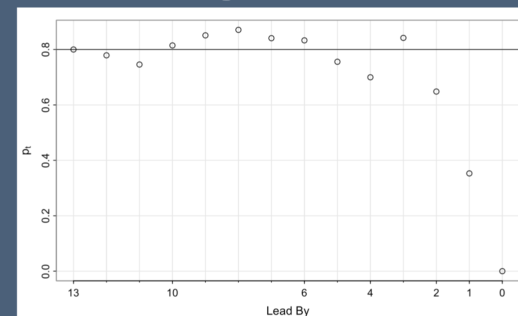


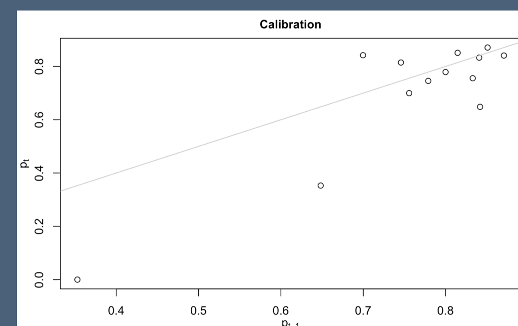


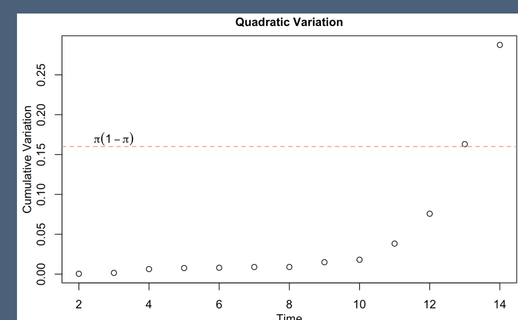
Example of Threshold Martingale

- AR(1) process
  - Points are values of X_n, X_{n+1}, ..., X_T
  - Threshold τ is 80% point of initial forecast distribution at time n+1
  - Shaded regions determine p_t









Hard to tell if one sequence is a martingale,
so,
suppose we have a lot of sequences!

Ensemble Diagnostic

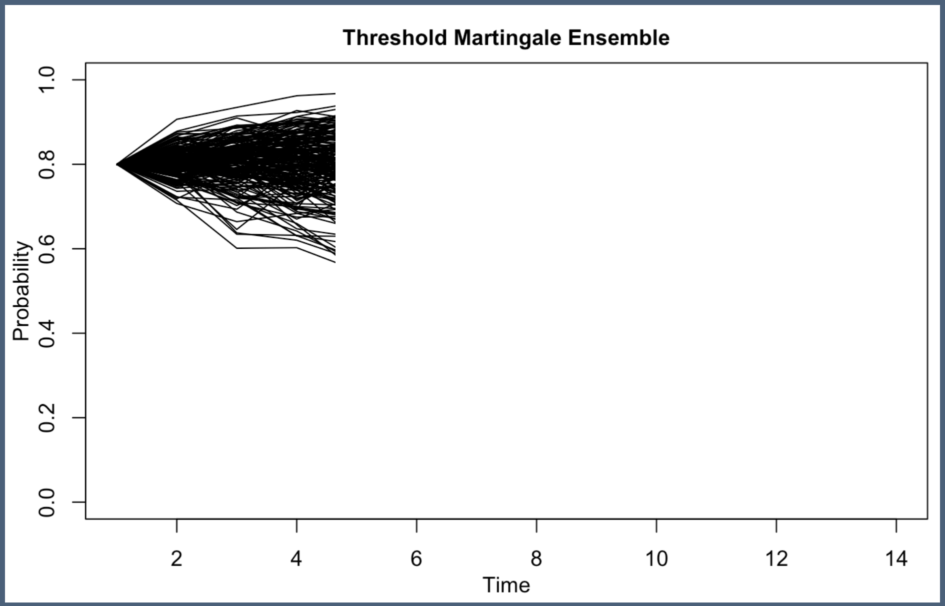
- Requirement of approach
  - Forecast provides a distribution
- Properties of approach
  - Don't need to known how you constructed distribution associated with forecast
    - – Guesswork
    - – ARMA process
    - – Neural network
  - Calculation is routine, fast
- On-line feedback
  - Don't have to wait until target shows up

Example: All is well

- Context
  - Collection of 250 time series forecasts
  - Target is T = n + 14
  - Forecasts working as they should be
- Threshold martingales

Lecture_24.Rmd

π = 0.80

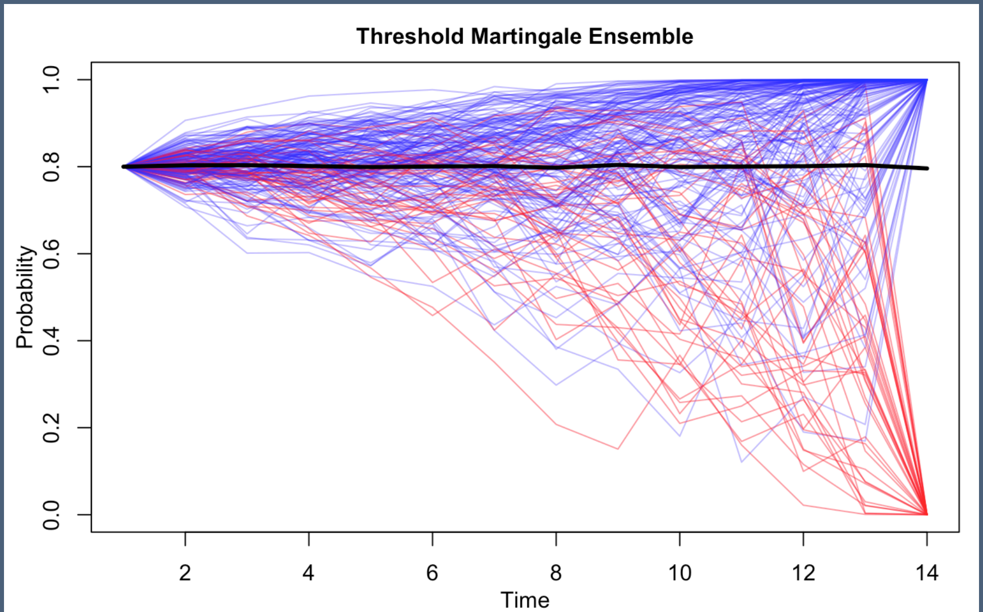


Observed sequence is on-line, revealed gradually as target date approaches

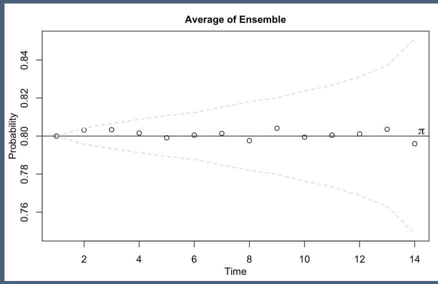
Example: All is well

- Context
  - Collection of 250 time series forecasts
  - Target is T = n + 14
  - Forecasts working as they should be
- Threshold martingales

Lecture_24.Rmd



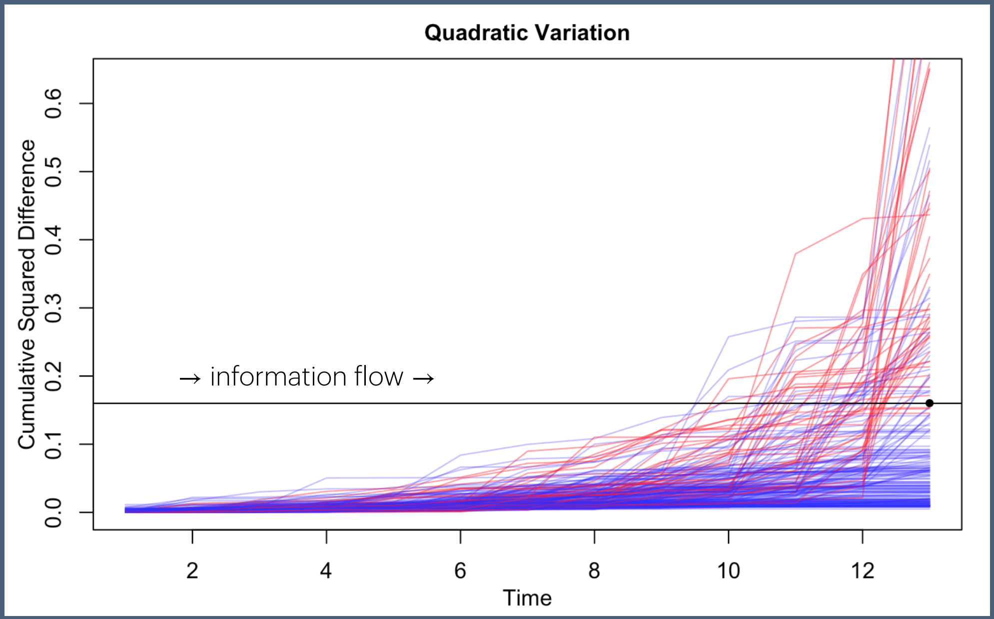
Colors added after the fact



Example: All is well

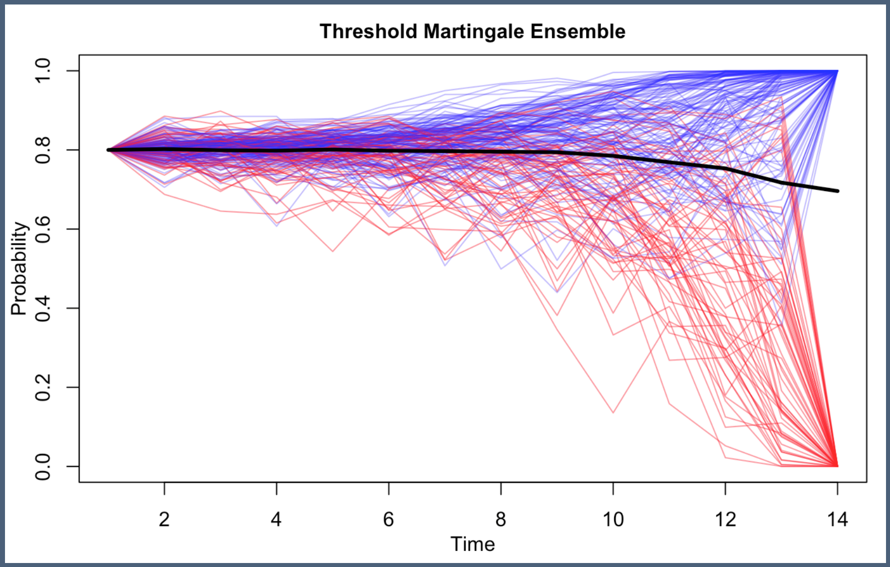
- Context
  - Collection of 250 time series forecasts
  - Target is T = n + 14
  - Forecasts working as they should be
- Quadratic variation

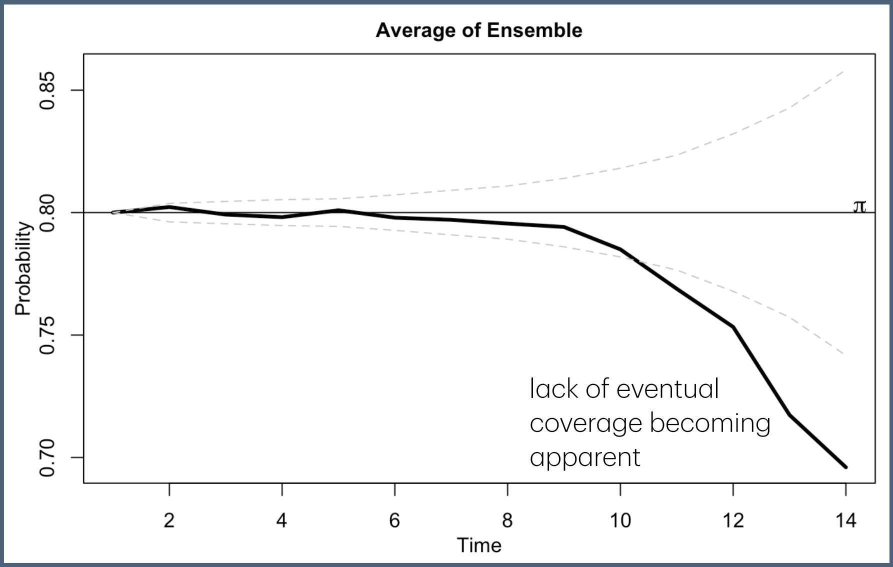
Lecture_24.Rmd



Example: All is NOT well, # 1

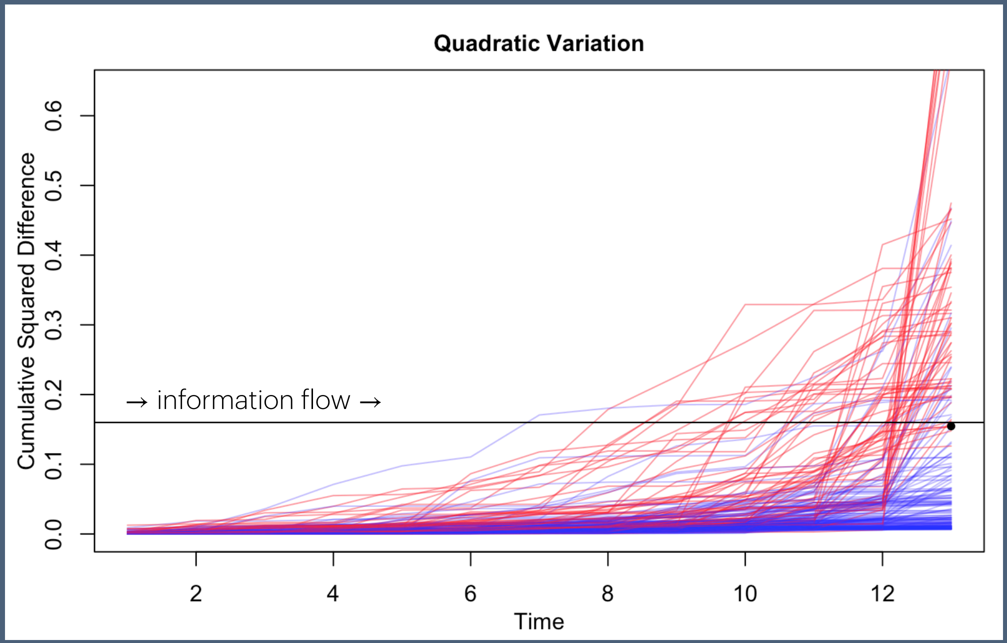
- Context
  - Collection of 250 time series forecasts
  - Target is T = n + 14
  - Forecasts NOT working as they should be: coefficient used to construct forecast moving around
- Threshold martingales





Example: All is NOT well #1

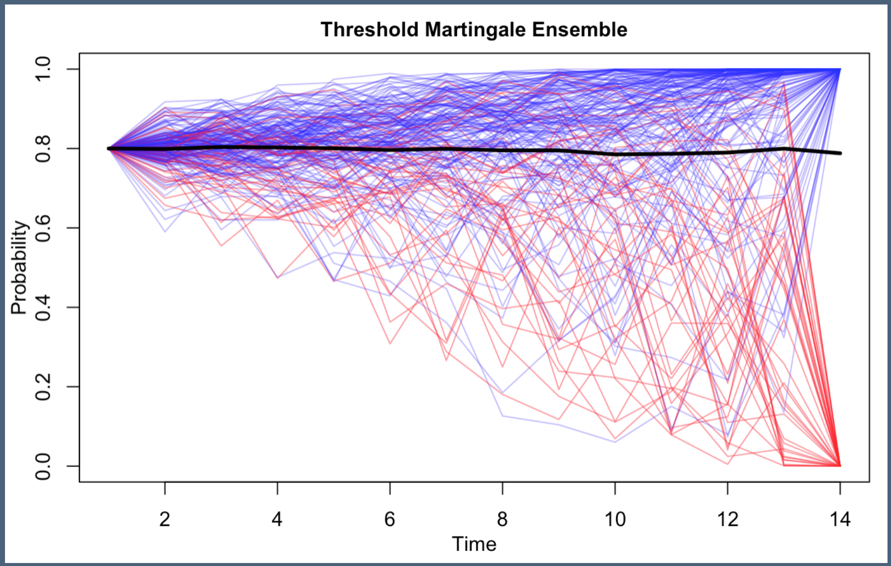
- Context
  - Collection of 250 time series forecasts
  - Target is T = n + 14
  - Forecasts working as they should be
- Quadratic variation

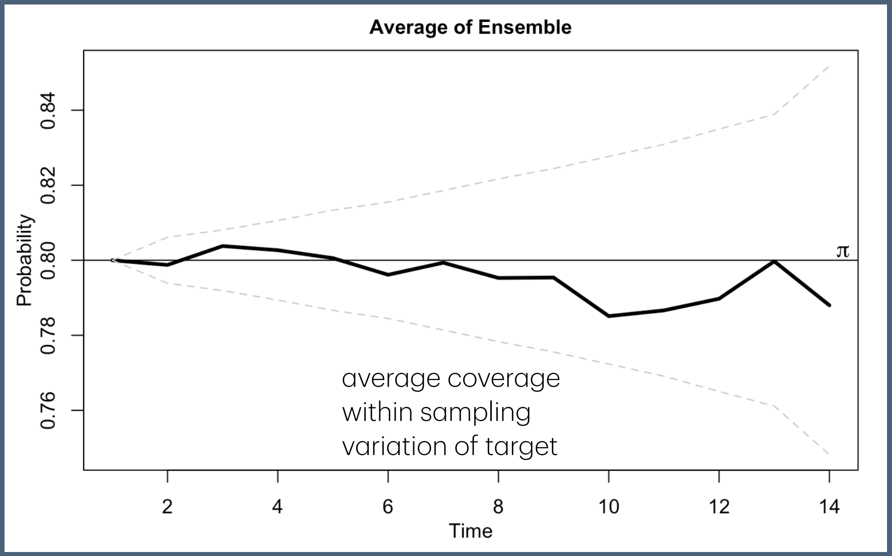


Looks OK in this simulation even though the process is not behaving correctly

Example: All is NOT well, #2

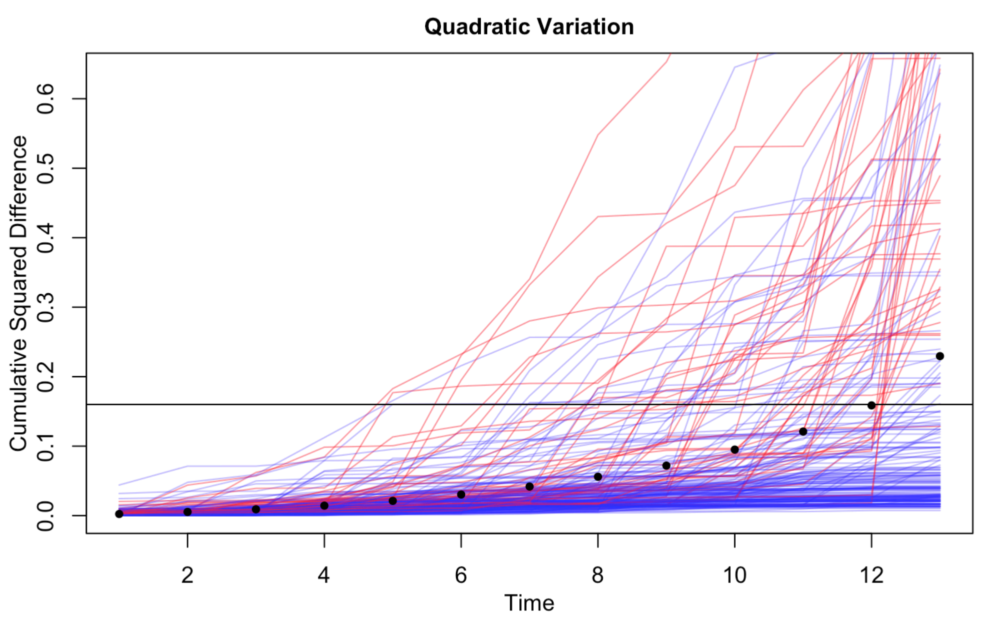
- Context
  - Collection of 250 time series forecasts
  - Target is T = n + 14
  - Forecasts NOT working as they should be: forecast “bouncing” around too much
- Threshold martingales





Example: All is NOT well

- Context
  - Collection of 250 time series forecasts
  - Target is T = n + 14
  - Forecasts working as they should be
- Quadratic variation



Excess volatility is apparent

What's next?

- ARCH/GARCH models
  - Asset values are hard to predict: predicting stock returns
  - Asset volatility is predictable: predicting volatility of stock returns