Conditional Mean Forecasts

- Under squared error loss, the optimal *unconditional* point forecast is the mean E*y*

$$\mathrm{E}(y) = \int y f(y) dy$$

- Under squared error loss, the optimal *conditional* point forecast is the *conditional mean*

$$\mathrm{E}(y \mid x) = \int y f(y \mid x) dy$$

where $f(y|x)$ is the conditional density of y given x

Conditional Variables

- The variable x is the information available for us to forecast y
- For example, to forecast the wage we let x be the person's sex, race and educational level.
- The unconditional mean $\mathrm{E}y$ is the average wage in the entire population
- The conditional mean $\mathrm{E}(y|x)$ is the average wage in the population for the subpopulation with the variables x.

Conditional Mean

- What is $E(y|x)$?
- When x is discrete, then $E(y|x)$ is simply the mean in the subpopulation.
- For example, the average wage among white male college graduates.
- When x is continuously distributed, this definition does not work.

Conditional Mean

- Let $f(y,x)$ denote the joint density of y and x. Then the conditional density of y given x is

$$f(y \mid x) = \frac{f(y,x)}{f(x)}$$

- It is the distribution of y holding x fixed.
- It is a slice of the joint density
- Then the conditional mean is

$$\mathrm{E}(y \mid x) = \int y f(y \mid x) dy$$

Forecast Intervals

- Forecast intervals are constructed from the conditional distribution $F(y|x)$.
- The endpoints are the conditional quantiles.
- Definition: The α'th conditional quantile of y given x is the number $q_\alpha(x)$ which satisfies

$$\alpha = F\left(q_\alpha(x) \mid x\right)$$

- It looks more complicated, but it is identical with the case with no x.

When does Conditioning Help?

- It helps if y and x are *dependent* or *correlated*.
- If y and x are independent, then $f(y|x)=f(y)$ and $E(y|x)=E(y)$.
  - There is no gain from conditioning.
- To optimally forecast unknown y we want to use observable variables x which are highly correlated with y.

Summary – Conditional Forecasts

- Information improves forecasts
- Conditioning on relevant variables reduces the risk (expected loss) of forecasts
- Point and interval forecasts are functions of the conditioning variables

Forecasting Economic Time Series

Economic forecasts rely on time series data – observations which are recorded sequentially over time.

Time Series Data

- A time series is written as y_t
- The index t denotes the time period.
- A time period may be a year, quarter, month, week, day, transaction, or any other time unit.
- We call this the data frequency.

Lags and Leads

- We will often talk about lags and leads
- The first lag of y_t is written $y_{t\text{-}1}$
  - It is the observation from the previous period
  - For example, the lag of November is October
- The second lag is y_{t-2}, the k'th lag is y_{t-k}
- The first lead of y_t is y_{t+1}
- The k'th lead is y_{t+k}

Time Series Samples

- A historical sample is a set of observations in contiguous time, written as

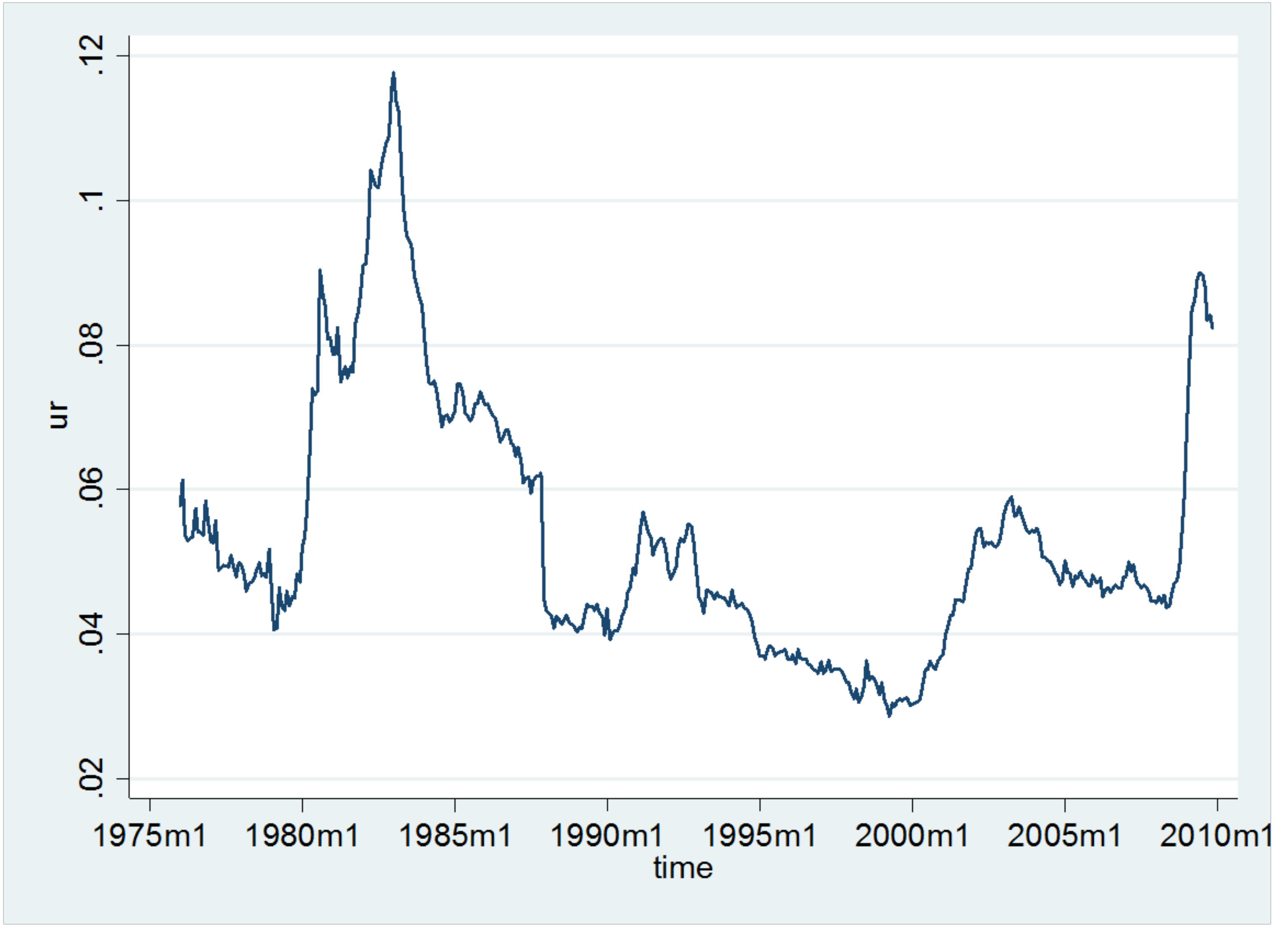
$$\{y_1, y_2, \ldots, y_T\}$$

- T is the number of observations in-sample.
- The number of observations does not equal the number of years, unless the frequency is annual

Examples

- Number of Unemployed
- Unemployment Rate
- GDP
- Real GDP
- Price Level
- Inflation Rate
- New Housing Starts

Wisconsin Unemployment Rate



Forecast Period

- In-sample observations: $\{y_1, y_2, ..., y_T\}$
- Out-of-sample period: $\{y_{T+1}, y_{T+2}, ..., y_{T+h}\}$
- h is called the forecast horizon

Forecast Notation

- We denoted $\hat{y}$ as the point forecast for y.
- This suggests $\hat{y}_{T+h}$ as the point forecast for y_{T+h}.
- But is not enough. While it is the forecast for the time series at time period $T+h$, it is not clear when the forecast is made.
  - At time period T
  - At time period $T+1$
  - At time period $T+h-1$

Notation

- We will use the notation

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

to refer to the forecast of y_{t+h} made at time *t*.

- Thus $\hat{y}_{T+h|T}$, $\hat{y}_{T+h|T+1}$, $\hat{y}_{T+h|T+2}$, etc. are the sequence of forecasts of y_{T+h} made in time periods *T*, *T+1*, *T+2*, etc.

Notation

- Similarly, the forecast distribution and density for y_{t+h} made at time t will be written as

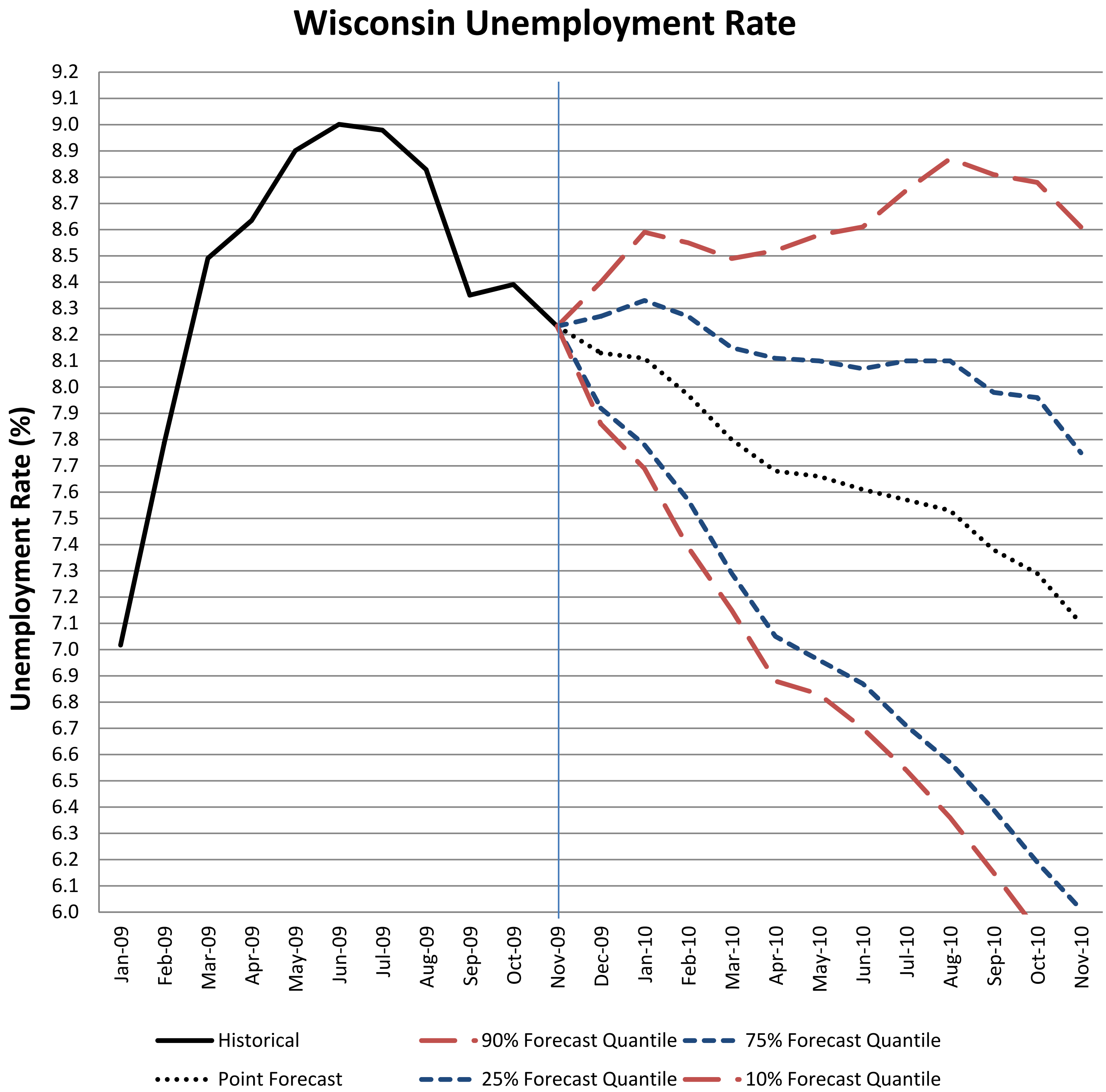
$$F_{t+h|t}(y)$$

and

$$f_{t+h|t}(y)$$

Extrapolative Forecasts and Fan Charts

- At time *T*, make a sequence of forecasts for time periods *T*+1, *T*+2, *T*+3,..., *T*+h
- Point forecasts: $\hat{y}_{T+1|\mathrm{T}}$, $\hat{y}_{T+2|\mathrm{T}}$, ..., $\hat{y}_{T+h|\mathrm{T}}$
- Add interval forecasts
- Plot over forecast horizon
- This is called a *fan chart*.
- The intervals tend to fan out with the forecast horizon.



Wisconsin Unemployment Rate
Unemployment Rate (%)
9.2
9.1
9.0
8.9
8.8
8.7
8.6
8.5
8.4
8.3
8.2
8.1
8.0
7.9
7.8
7.7
7.6
7.5
7.4
7.3
7.2
7.1
7.0
6.9
6.8
6.7
6.6
6.5
6.4
6.3
6.2
6.1
6.0
Jan-09
Feb-09
Mar-09
Apr-09
May-09
Jun-09
Jul-09
Aug-09
Sep-09
Oct-09
Nov-09
Dec-09
Jan-10
Feb-10
Mar-10
Apr-10
May-10
Jun-10
Jul-10
Aug-10
Sep-10
Oct-10
Nov-10
Historical
90% Forecast Quantile
75% Forecast Quantile
Point Forecast
25% Forecast Quantile
10% Forecast Quantile

Extrapolative Forecasts

	Point Forecast	50% Interval Forecast	80% Interval Forecast
2009:12	8.1%	(7.9%, 8.3%)	(7.9%, 8.4%)
2010: 1	8.1%	(7.8%, 8.3%)	(7.7%, 8.6%)
2010: 2	8.0%	(7.6%, 8.3%)	(7.4%, 8.6%)
2010: 3	7.8%	(7.3%, 8.2%)	(7.2%, 8.5%)
2010: 4	7.7%	(7.1%, 8.1%)	(6.9%, 8.5%)
2010: 5	7.7%	(7.0%, 8.1%)	(6.8%, 8.6%)
2010: 6	7.6%	(6.9%, 8.1%)	(6.7%, 8.6%)
2010: 7	7.6%	(6.7%, 8.1%)	(6.5%, 8.8%)
2010: 8	7.5%	(6.6%, 8.1%)	(6.4%, 8.9%)
2010: 9	7.4%	(6.4%, 8.0%)	(6.2%, 8.8%)
2010:10	7.3%	(6.2%, 8.0%)	(5.9%, 8.8%)
2010:11	7.1%	(6.0%, 7.8%)	(5.7%, 8.6%)

Information Set

- To forecast y_{t+h} at time t we use relevant information.
- Most information is values of other economic variables.
- Those observed up to time t.
- This includes previous values of the variable y_t.
- It can also include other relevant variables.
- All this information is the *Information Set*, written as Ω_t.
- For example $\Omega_t = \{y_1, y_2, y_3, \ldots, y_t\}$ is the set of previous values.
- $\Omega_t = \{y_1, x_1, y_2, x_2, y_3, x_3, \ldots, y_t, x_t\}$ includes another variable x_t

Forecast from the Information Set

- The conditional density of y_{t+1} given Ω_t is

$$f\left(y_{t+1} \mid \Omega_t\right) = \frac{f\left(y_{t+1}, \Omega_t\right)}{f\left(\Omega_t\right)}$$

- This is the (conditional) probability density, given the information Ω_t.
- The conditional distribution is its integral

$$F\left(y \mid \Omega_t\right) = \int_{-\infty}^{y} f\left(u \mid \Omega_t\right) du$$

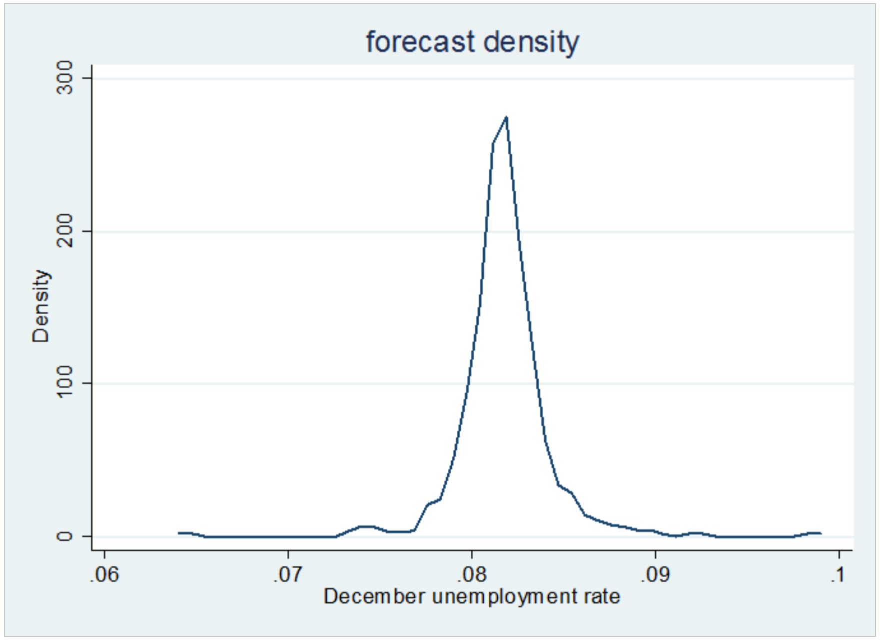
Forecasts

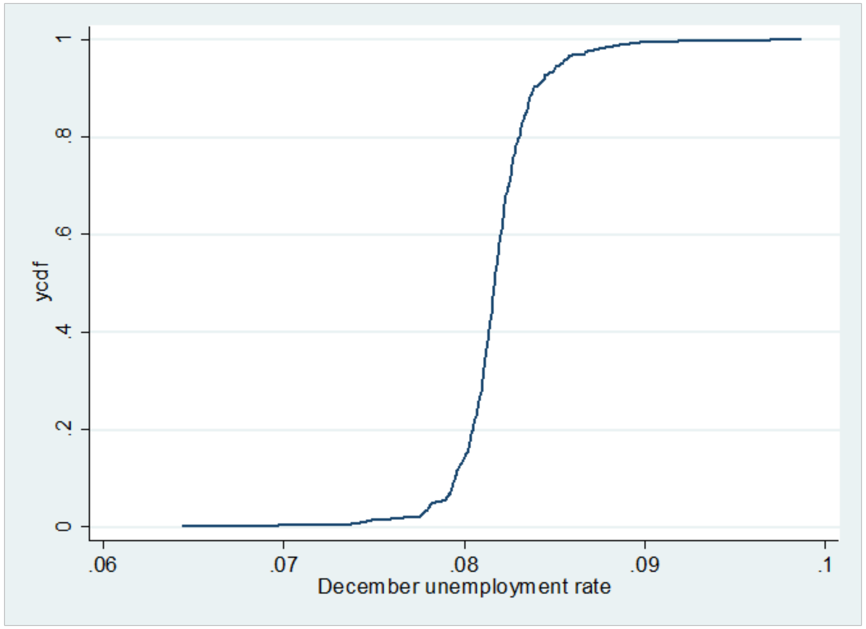
- The point forecast is the conditional mean

$$\mathrm{E}\left(y_{t+h} \mid \Omega_t\right) = \int y f\left(y \mid \Omega_t\right) dy$$

- Interval forecasts are quantiles of the conditional distribution function $F(y|\Omega_t)$

Forecast Density and Distribution





Actual Forecasting

- Even if the variables in the information set Ω_t are known, the conditional mean function $E(y_{t+h} \mid \Omega_t)$ is unknown
  - The functional form is unknown
  - The parameters of the function are unknown
- Thus to make an actual forecast, we need to:
  - Create an approximate model for $E(y_{t+h} \mid \Omega_t)$
  - Estimate the model parameters from data.

Diebold’s Six Considerations

1. Decision Environment
   - Do you have a specialized loss function?
2. Forecast Object
   - A time series
   - An event (a binary time series)
     - In this case the forecast should be a *probability*
   - A duration or timing

3. Forecast Statement
   – Is the goal point estimation alone?
   – Do you want an interval or density forecast?
4. Forecast Horizon
   – 1-step-ahead
   – *h*-step-ahead
   – *h*-step-ahead extrapolation (1 through *h*)
5. The Information Set
   – Univariate (past history of series alone)
   – Multivariate (multiple series)

6. Methods and Complexity
   - How do you approximate the unknown functions?
   - Do you use a mechanical forecasting rule (e.g. exponential smoothing) or a statistical model?
   - How do you estimate the parameters?
   - To keep from overfitting
     - Enforce simplicity
     - Parsimony principle: Fewer parameters is better
     - Shrinkage principle: Coax your model towards simplicity
     - KISS principle: Keep It Sophisticatedly Simple