Lecture 4: Updating and Forecasting with New Information

UCSD, January 30, 2017

Allan Timmermann[1]

[1]UC San Diego

Updating and forecasting

- A good forecasting model closely tracks how data evolve over time
- Important to update beliefs about the predicted variable or the forecasting model as new information arrives
  - Bayes rule
- Suppose the current "state" is unobserved. For example we may not know in real time if the economy is in a recession
  - Filtering
  - Nowcasting: obtaining the best estimate of the current state
- How do we accurately and efficiently update our forecasts as new information arrives?
  - Kalman filter (continuous state)
  - Regime switching model (small number of states)

Bayes' rule

- Bayes rule for two random variables A and B:

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

- $P(A)$: probability of event A
- $P(B)$: probability of event B
- $P(B|A)$: probability of event B given that event A occurred

Bayes' rule (cont.)

- Let θ be some unknown model parameters while y are observed data. From Bayes' rule (setting $\theta = B$, $y = A$)

$$P(\theta|y) = \frac{P(y|\theta)P(\theta)}{P(y)}$$

- Given some data, y, what do we know about model parameters θ?
- If we are only interested in θ, we can ignore $p(y)$:

$$\underbrace{P(\theta|y)}_{posterior} \propto \underbrace{P(y|\theta)}_{likelihood}\underbrace{P(\theta)}_{prior}$$

- We start with **prior** beliefs before seeing the data. Updating these beliefs with the observed data, we get **posterior** beliefs
  - Parameters, θ, that do not fit the data become less likely
  - For example, if θ_H is 'high mean returns' and we observe a data sample with low mean returns, then we put less weight on θ_H

Examples of Bayes' rule

- B : European recession. A : European growth rate of -1%

$$P(recession|g=-1\%) = \frac{P(g=-1\%|recession)P(recession)}{P(g=-1\%)}$$

- B : bear market. A : negative returns of -5%

$$P(bear|r=-5\%) = \frac{P(r=-5\%|bear)P(bear)}{P(r=-5\%)}$$

- Here $P\left(recession\right)$ and $P(bear)$ are the initial probabilities of being in a recession/bear market (before observing the data)

Understanding updating: bivariate normal distribution

- Suppose two random variables Y and X are normally distributed

$$\begin{pmatrix} Y \\ X \end{pmatrix} = N\left(\begin{pmatrix} \mu_y \\ \mu_x \end{pmatrix}, \begin{pmatrix} \sigma_y^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_x^2 \end{pmatrix}\right)$$

- μ_y and μ_x are the initial (unconditional) expected values of y and x
- $\sigma_y^2, \sigma_x^2, \sigma_{xy}$ are variances and covariance
- **Conditional** mean and variance of Y given an observation $X = x$:

$$\begin{aligned} E[Y|X = x] &= \mu_y + \frac{\sigma_{xy}}{\sigma_x^2}(x - \mu_x) \\ Var(Y|X &= x) = \sigma_y^2 - \frac{\sigma_{xy}^2}{\sigma_x^2} \end{aligned}$$

- If Y and X are positively correlated ($\sigma_{xy} > 0$) and we observe a high value of X ($x > \mu_x$), then we increase our expectation of Y
- Just like a linear regression: σ_{xy}/σ_x^2 is the beta coefficient

Kalman Filter: Background

- The Kalman filter is an algorithm for linear updating and prediction
- Introduced by Kalman in 1960 for engineering applications
- Method has found great use in many disciplines, including economics and finance
- Kalman Filter gives an updating rule that can be used to revise our beliefs as we see more and more data
- For models with normally distributed variables, the filter can be used to write down the likelihood function

Kalman Filter (Wikipedia) I

- "Kalman filtering, also known as linear quadratic estimation (LQE), is an algorithm that uses a series of measurements observed over time, containing noise (random variations) and other inaccuracies, and produces estimates of unknown variables that tend to be more precise than those based on a single measurement alone. More formally, the Kalman filter operates recursively on streams of noisy input data to produce a statistically optimal estimate of the underlying system state. The filter is named for Rudolf (Rudy) E. Kálmán, one of the primary developers of its theory.
- The Kalman filter has numerous applications in technology. A common application is for guidance, navigation and control of vehicles, particularly aircraft and spacecraft. Furthermore, the Kalman filter is a widely applied concept in time series analysis used in fields such as signal processing and econometrics.

Kalman Filter (Wikipedia) II

- The algorithm works in a two-step process. In the **prediction step**, the Kalman filter produces estimates of the current state variables, along with their uncertainties. Once the outcome of the next measurement (necessarily corrupted with some amount of error, including random noise) is observed, these estimates are **updated** using a weighted average, with more weight being given to estimates with higher certainty. Because of the algorithm's recursive nature, it can run in real time using only the present input measurements and the previously calculated state and its uncertainty matrix; no additional past information is required."

Kalman Filter: Models in state space form

- Let S_t be an unobserved (state) variable while y_t is an observed variable. A model that shows how y_t is related to S_t and how S_t evolves is called a **state space model**. This has two equations:
  - State equation (unobserved/latent):

$$S_t = \phi \times S_{t-1} + \varepsilon_{st}, \ \varepsilon_{st} \sim (0, \sigma_s^2) \tag{1}$$

  - Measurement equation (observed)

$$y_t = B \times S_t + \varepsilon_{yt}, \ \varepsilon_{yt} \sim (0, \sigma_y^2) \tag{2}$$

  - Innovations are uncorrelated with each other:

$$Cov(\varepsilon_{st}, \varepsilon_{yt}) = 0$$

Example 1: AR(1) model in state space form

- AR(1) model

$$y_t = \phi y_{t-1} + \varepsilon_t$$

- This can be written in state space form as

$$\begin{aligned} S_t &= \phi S_{t-1} + \varepsilon_t && \text{state eq.} \\ y_t &= S_t && \text{measurement eq.} \end{aligned}$$

- with $B = 1, \sigma_s^2 = \sigma_\varepsilon^2$, and $\sigma_y^2 = 0$
  - very simple: no error in the measurement equation: y_t is observed without error

Example 2: MA(1) model in state space form

- MA(1) model with unobserved shocks ε_t :

$$y_t = \varepsilon_t + \theta\varepsilon_{t-1}$$

- This can be written in state space form

$$\begin{pmatrix} S_{1t} \\ S_{2t} \end{pmatrix} = \underbrace{\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}}_{\phi} \begin{pmatrix} S_{1t-1} \\ S_{2t-1} \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \end{pmatrix} \varepsilon_t$$

$$y_t = \underbrace{\begin{pmatrix} 1 & \theta \end{pmatrix}}_{B} \begin{pmatrix} S_{1t} \\ S_{2t} \end{pmatrix} = \varepsilon_t + \theta\varepsilon_{t-1}$$

- Note that $S_{1t} = \varepsilon_t, S_{2t} = S_{1t-1} = \varepsilon_{t-1}$

Example 3: Unobserved components model

- The unobserved components model consists of two equations

$$\begin{aligned} y_t &= S_t + \varepsilon_{yt} \qquad (B = 1) \\ S_t &= S_{t-1} + \varepsilon_{st} \qquad (\phi = 1) \end{aligned}$$

- y_t is observed with noise
- S_t is the underlying "mean" of y_t. This is smoother than y_t
- This model can be written as an ARIMA(0,1,1):

$$\begin{aligned} \Delta y_t &\equiv y_t - y_{t-1} = S_t - S_{t-1} + \varepsilon_{yt} - \varepsilon_{yt-1} \\ &= \varepsilon_{s_t} + \varepsilon_{yt} - \varepsilon_{yt-1} : \quad MA(1) \\ Cov(\Delta y_t, \Delta y_{t-j}) &= -\sigma_y^2 \text{ for } j = 1, \; = 0 \text{ for } j \geq 2 \end{aligned}$$

Kalman Filter: Advantages

- The state equation in (1) is in AR(1) form and so is easy to iterate forward. The $h-$step-ahead forecast of the state given its current value, S_t, is given by

$$E_t[S_{t+h}|S_t] = \phi^h S_t$$

- In practice we don't observe S_t and so need an estimate of this given current information, $S_{t|t}$, or past information, $S_{t|t-1}$
- Updating the Kalman filter through newly arrived information is easy

Kalman Filter Updating Equations

- $\mathcal{I}_t = \{y_t, y_{t-1}, y_{t-2}, ...\}$. Current information
- $\mathcal{I}_{t-1} = \{y_{t-1}, y_{t-2}, y_{t-3}, ...\}$. Lagged information
- y_t : random variable we want to predict
- $y_{t|t-1}$: best prediction of y_t given information at $t-1$, $\mathcal{I}_{t-1}$
- $S_{t|t-1}$: best prediction of S_t given information at $t-1$, $\mathcal{I}_{t-1}$
- $S_{t|t}$: best "prediction" (or nowcast) of S_t given information $\mathcal{I}_t$
- Define mean squared error (MSE) values associated with the forecasts of S_t and y_t

$$\begin{aligned} MSE^S_{t|t-1} &= E[(S_t - S_{t|t-1})^2] \\ MSE^S_{t|t} &= E[(S_t - S_{t|t})^2] \\ MSE^y_{t|t-1} &= E[(y_t - y_{t|t-1})^2] \end{aligned}$$

Prediction and Updating Equations

- Using the state, measurement, and MSE equations, the Kalman filter gives a set of **prediction equations**:

$$\begin{aligned} S_{t|t-1} &= \phi S_{t-1|t-1} \\ MSE^{S}_{t|t-1} &= \phi^2 MSE^{S}_{t-1|t-1} + \sigma^2_s \\ y_{t|t-1} &= B \times S_{t|t-1} \\ MSE^{y}_{t|t-1} &= B^2 \times MSE^{S}_{t|t-1} + \sigma^2_y \end{aligned}$$

- Similarly, we have a pair of **updating equations** for S:

$$\begin{aligned} S_{t|t} &= S_{t|t-1} + B \left(MSE^{S}_{t|t-1} / MSE^{y}_{t|t-1} \right) (y_t - y_{t|t-1}) \\ MSE^{S}_{t|t} &= MSE^{S}_{t|t-1} \left[1 - B^2 \left(MSE^{S}_{t|t-1} / MSE^{y}_{t|t-1} \right) \right] \end{aligned}$$

Prediction and Updating Equations

- Intuition for updating equations $(B = 1)$

$$S_{t|t} = S_{t|t-1} + \frac{MSE^S_{t|t-1}}{MSE^y_{t|t-1}}(y_t - y_{t|t-1})$$

- $S_{t|t}$: estimate of current (t) state given current information $\mathcal{I}_t$
- $S_{t|t-1}$: old $(t-1)$ estimate of state S_t given $\mathcal{I}_{t-1}$
- $MSE^S_{t|t-1} / MSE^y_{t|t-1}$: amount by which we update our estimate of the current state after we observe y_t. This is small if $MSE^y_{t|t-1}$ is big (noisy data) relative to $MSE^S_{t|t-1}$, i.e., $\sigma^2_y >> \sigma^2_s$
- $(y_t - y_{t|t-1})$: surprise (news) about y_t
- If y_t is higher than we expected, $(y_t - y_{t|t-1}) > 0$, we increase our expectations about the state: $S_{t|t} > S_{t|t-1}$. The updating equation tells us by how much

Starting the Algorithm

- At $t = 0$, we have not observed any data, so we must make our best guesses of $S_{1|0}$ and $MSE^S_{1|0}$ without data by picking a pair of initial conditions. This gives $y_{1|0}$ and $MSE^y_{1|0}$ from the prediction equations
- At $t = 1$ we observe y_1. The updating equations generate $S_{1|1}$ and $MSE^s_{1|1}$. The prediction equations then generate forecasts for the second period
- At $t = 2$ we observe y_2, and the cycle continues to give sequences of predictions of the states, $\{S_{t|t}\}$ and $\{S_{t|t-1}\}$
- Keep on iterating to get a sequence of estimates

Filtered versus smoothed states

- $S_{t|t}$: **filtered states**: estimate of the state at time t given information up to time t
  - Uses only historical information
  - What is my best guess of S_t given my current information?
  - "Filters" past historical information for noise
- $S_{t|T}$: **smoothed states**: estimate of the state at time t given information up to time T
  - Uses the full sample up to time $T \geq t$
  - Less sensitive to noise and thus tends to be smoother than the filtered states
  - Information on y_{t-1}, y_t, y_{t+1} help us more precisely estimate the state at time t, S_t

Practical applications of the Kalman filter

- Common to use Kalman filter to estimate adaptive forecasting models with time-varying relations:

$$y_{t+1} = \beta_t x_t + \varepsilon_{t+1}$$

- y_t, x_t : observed variables
- β_t : time-varying coefficient (unobserved state variable)
- Alternative specifications for β_t :

$$\begin{aligned} \beta_t - \bar{\beta} &= \phi(\beta_{t-1} - \bar{\beta}) + u_t : \text{ mean-reverts to } \bar{\beta} \\ \beta_t &= \beta_{t-1} + u_t : \text{ random walk} \end{aligned}$$

Kalman filter in matlab

- Matlab has a Kalman filter called *ssm* (state space model). The setup is

$$
\begin{aligned}
x_t &= A_t x_{t-1} + B_t u_t \\
y_t &= C_t x_t + D_t e_t
\end{aligned}
$$

- x_t : unobserved state (our S_t)
- y_t : observed variable
- u_t, e_t : uncorrelated noise processes with variance of one
- model = ssm(A,B,C,D,'StateType',stateType); % state space model
- modelEstimate = estimate(model,variable,params0,'lb',[0; 0])
- filtered = filter(modelEstimate,variable)
- smoothed = smooth(modelEstimate,variable)

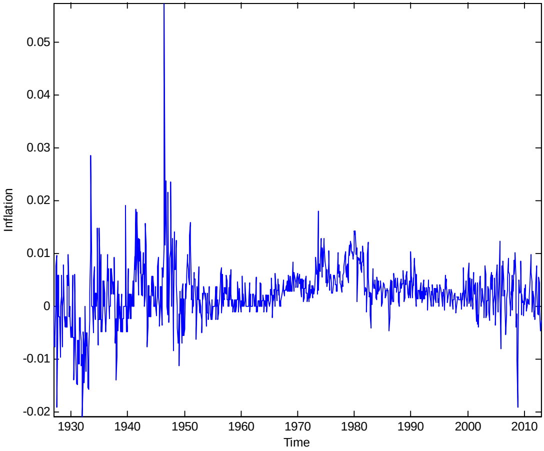
Kalman filter example: monthly inflation

- Unobserved components model for inflation

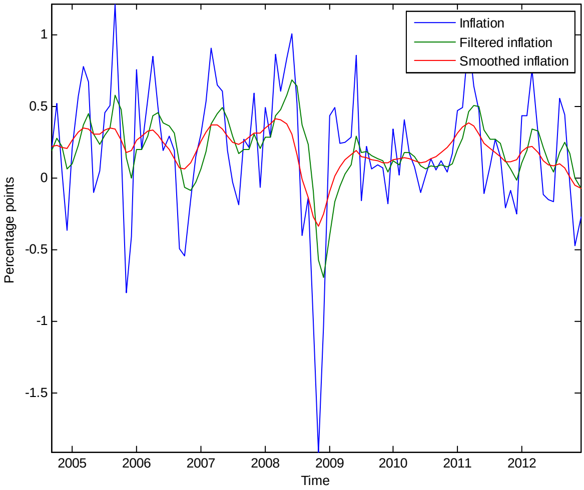
$$
\begin{aligned}
x_t &= x_{t-1} + \sigma_u u_t \\
y_t &= x_t + \sigma_e e_t
\end{aligned}
$$

- $A = 1$; % state-transition matrix ($A = \phi$ in our notation)
- $B = NaN$; % state-disturbance-loading matrix ($B = \sigma_S$)
- $C = 1$; % measurement-sensitivity matrix ($C = B$ in our notation)
- $D = NaN$; % observation-innovation matrix ($D = \sigma_y$)
- stateType $= 2$; % sets state equation to be a random walk

Application of Kalman filter to monthly US inflation



Kalman filter estimates of inflation (last 100 obs.)



Kalman filter take-aways

- Kalman filter is a very popular approach for dynamically updating linear forecasts
- Used to estimate ARMA models
- Used throughout engineering and the social sciences
- Fast, easy algorithm
- Optimal updating equations for normally distributed data

Nowcasting

- Nowcasting refers to "estimating the present"
- Nowcasting extracts information about the present state of some variable or system of variables
  - distinct from traditional forecasting
- Nowcasting only makes sense if the present state is unknown—otherwise nowcasting would just amount to checking the current value
- Example: Use a single unobserved state variable to summarize the state of the economy, e.g., the daily point in the business cycle
- Variables such as GDP are actually observed with large measurement errors (revisions)

Jagged edge data

- Macroeconomic data such as GDP, monetary aggregates, consumption, unemployment figures or housing starts as well as financial data extracted from balance sheets and income statements are published infrequently and sometimes at irregular intervals
- Delays in the publication of macro variables differ across variables
- Irregular data releases (release date changes from month to month) generate what is often called "jagged edge" data
- A forecaster can only use the data that is available on any given date and needs to pay careful attention to which variables are in the information set

Aruoba, Diebold and Scotti daily business cycle indicator

- ADS model the daily business cycle, S_t, as an unobserved variable that follows a (zero-mean) AR(1) process:

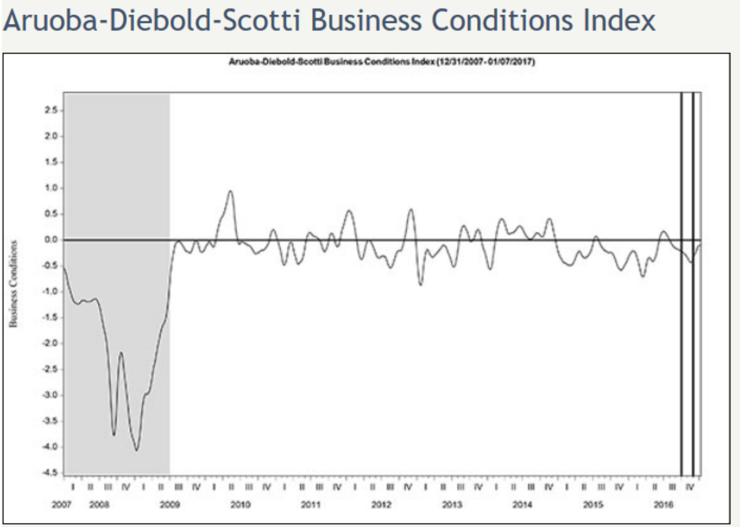
$$S_t = \phi S_{t-1} + \epsilon_t$$

- Although S_t is unobserved, we can extract information about it from its relation with a set of observed economic variables $y_{1t}, y_{2t}, ...$
- At the daily horizon these variables follow processes:

$$y_{it} = k_i + \beta_i S_t + \gamma_i y_{i,t-D_i} + u_{it}, \quad i = 1, .., n$$

- D_i equals seven days if the variable is observed weekly, etc.

Aruoba, Diebold and Scotti index from Philly Fed



ADS four-variable model

- The ADS model can be written in state-space form
- For example, a model could use the following observables:
  - initial jobless claims (weekly, y_{1t})
  - personal income (monthly, y_{2t})
  - industrial production (monthly, y_{3t})
  - GDP (quarterly, y_{4t})
- Kalman filter can be used to extract and update estimates of the unobserved common variable that tracks the state of the economy
- Kalman filter is well suited for handling missing data
- If all elements of y_t are missing on a given day, we skip the updating step

Automatic time series forecasting

- Hyndman and Khandakar (2008) develop a forecast package for *R* using exponential smoothing for forecasting
  - http://CRAN.R-project.org/package=forecasting
- Holt-Winters additive method:

$$
\begin{array}{ll}
\text{Level} & l_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(l_{t-1} + b_{t-1}) \\
\text{Growth} & b_t = \beta^*(l_t - l_{t-1})(1-\beta^*)b_{t-1} \\
\text{Seasonal} & s_t = \gamma(y_t - l_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m} \\
\text{Forecast} & f_{t+h|t} = l_t + b_t h + s_{t-m+h_m^+}
\end{array}
$$

- m : length of seasonality
- l_t : level of the series
- b_t : growth
- s_t : seasonal component
- α, β^*, γ : smoothing parameters
- Values of $l_0, b_0, s_{1-m}, ..., s_0, \alpha, \beta^*, \gamma$ are estimated from the data

Automatic time series forecasting

- Exponential smoothing methods can allow for additive or multiplicative errors
- Estimation uses maximum likelihood methods
- Model selection proceeds using the AIC
- Exponential smoothing and ARIMA models overlap:
    - Linear exponential smoothing models are special cases of ARIMA models
    - Nonlinear exponential smoothing models do not have an ARIMA representation

Markov Chains: Basics

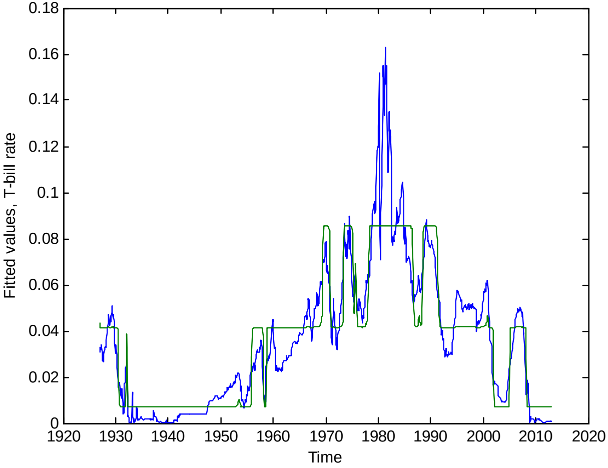
- Updating equations simplify a great deal if we only have two states, states 1 and 2, and want to know which state we are currently in
  - recession/expansion
  - inflation/deflation
  - bull/bear market
  - high volatility/low volatility

Markov Chains: Basics

- A first order (constant) Markov chain, S_t, is a random process that takes integer values $\{1, 2, ..., K\}$ with state transitions that depend only on the most recent state, S_{t-1}
- Probability of moving from state i at time $t-1$ to state j at time t is p_{ij} :

$$\begin{aligned} P(S_t &= j | S_{t-1} = i) = p_{ij} \\ 0 &\leq p_{ij} \leq 1 \\ \sum_{j=1}^{K} p_{ij} &= 1 \end{aligned}$$

Fitted values, 3-state model for monthly T-bill rates



Two-state Markov Chain

- With $K = 2$ states, the transition probabilities can be collected in a 2×2 matrix

$$\begin{aligned} P &= P(S_{t+1} = j | S_t = i) \\ &= \begin{bmatrix} p_{11} & p_{21} \\ p_{12} & p_{22} \end{bmatrix} = \begin{bmatrix} p_{11} & 1 - p_{22} \\ 1 - p_{11} & p_{22} \end{bmatrix} \end{aligned}$$

- Law of total probability: $p_{i1} + p_{i2} = 1$: we either stay in state i or we leave to state j
- p_{ii} : "stayer" probability - measure of state i's persistence

Basic regime switching model

- Simple two-state regime-switching model

$$\begin{aligned} y_{t+1} &= \mu_{s_{t+1}} + \sigma_{s_{t+1}} \varepsilon_{t+1}, \;\; \varepsilon_{t+1} \sim N(0,1) \\ P(s_{t+1} &= j | s_t = i) = p_{ij} \end{aligned}$$

- $\mu_{s_{t+1}}$: mean in state s_{t+1}
- $\sigma_{s_{t+1}}$: volatility in state s_{t+1}
- s_{t+1} matters for both the mean and volatility of y_{t+1}

$$\begin{aligned} \text{if } s_{t+1} &= 1 : y_{t+1} = \mu_1 + \sigma_1 \varepsilon_{t+1} \\ \text{if } s_{t+1} &= 2 : y_{t+1} = \mu_2 + \sigma_2 \varepsilon_{t+1} \end{aligned}$$

Updating state probabilities

- To predict y_{t+1}, we need to predict s_{t+1}. This depends on the current state, s_t
- Let $p_{1t|t} = prob(s_t = 1|\mathcal{I}_t)$ be the **current** probability of being in state 1 given all information up to time $t, \mathcal{I}_t$
  - If $p_{1t|t} = 1$, we know for sure that we are in state 1 at time t
  - Typically $p_{1t|t} < 1$ and there is uncertainty about the present state
- Let $p_{1t+1|t} = prob(s_{t+1} = 1|\mathcal{I}_t)$ be the **predicted** probability of being in state 1 next period $(t + 1)$, given $\mathcal{I}_t$

Updating state probabilities

- To be in state 1 at time $t+1$, we must have come from either state 1 or from state 2:

$$\begin{aligned} p_{1t+1|t} &= p_{11} \times p_{1t|t} + (1-p_{22}) \times p_{2t|t} \\ p_{2t+1|t} &= (1-p_{11}) \times p_{1t|t} + p_{22} \times p_{2t|t} \end{aligned}$$

- If $p_{1t|t} = 1$, we know for sure that we are in state 1 at time t. Then the equations simplify to

$$\begin{aligned} p_{1t+1|t} &= p_{11} \times 1 + (1-p_{22}) \times 0 = p_{11} \\ p_{2t+1|t} &= (1-p_{11}) \times 1 + p_{22} \times 0 = 1 - p_{11} \end{aligned}$$

Updating with two states

- Let $P(s_t = 1|y_{t-1})$ and $P(s_t = 1|y_{t-1})$ be our initial estimates of being in states 1 and 2 given information at time $t-1$
- In period t we observe a new data point: y_t
- If we are in state 1, the likelihood of observing y_t is $P(y_t|s_t = 1)$
- If we are in state 2, the likelihood of y_t is $P(y_t|s_t = 2)$
- If these are normally distributed, we have

$$
\begin{aligned}
P(y_t|s_t &= 1) = \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(\frac{-(y_t - \mu_1)^2}{2\sigma_1^2}\right) \\
P(y_t|s_t &= 2) = \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left(\frac{-(y_t - \mu_2)^2}{2\sigma_2^2}\right) \qquad (3)
\end{aligned}
$$

Bayesian updating with two states: examples I

- Use Bayes' rule to compute the updated state probabilities:

$$P(s_t = 1|y_t) = \frac{P(y_t|s_t = 1)P(s_t = 1)}{P(y_t)}, \quad \text{where}$$
$$P(y_t) = P(y_t|s_t = 1)P(s_t = 1) + P(y_t|s_t = 2)P(s_t = 2)$$

- Similarly

$$P(s_t = 2|y_t) = \frac{P(y_t|s_t = 2)P(s_t = 2)}{P(y_t)}$$

- Suppose that $\mu_1 < 0, \sigma_1^2$ is "large" so state 1 is a high volatility state with negative mean, while $\mu_2 > 0$ with small σ_2^2 so state 2 is a "normal" state

Bayesian updating with two states: examples II

- If we see a **large negative** y_t, this is most likely drawn from state 1 and so $P(y_t|s_t = 1) > P(y_t|s_t = 2)$. Then we revise upward the probability that we are currently (at time t) in state 1
- Example:
  - $\mu_1 = -3, \sigma_1 = 5, \mu_2 = 1, \sigma_2 = 2$
  - $P(s_t = 1|y_{t-1}) = 0.70,\ P(s_t = 2|y_{t-1}) = 0.30$: initial estimates
  - $p_{11} = 0.8, p_{22} = 0.9$
- Suppose we observe $y_t = -4$. Then from (3)

$$
\begin{aligned}
p(y_t|s_t &= 1) = Normpdf(-1/5) = 0.0782 \\
p(y_t|s_t &= 2) = Normpdf(-5/2) = 0.0088 \\
P(s_t &= 1|y_t) = \frac{0.0782 \times 0.70}{0.0782 \times 0.70 + 0.0088 \times 0.30} = 0.954 \\
P(s_t &= 2|y_t) = \frac{0.0088 \times 0.30}{0.0782 \times 0.70 + 0.0088 \times 0.30} = 0.046
\end{aligned}
$$

Bayesian updating with two states: examples III

- Because the observed value (-4%) is far more likely to have been drawn from state 1 than from state 2, we revise upwards our beliefs that we are **currently** in the first state from 70% to 95.4%
- Using p_{11} and p_{22}, our **forecast** of being in state 1 **next** period (at time $t+1$) is

$$P(s_{t+1} = 1 | y_t) = 0.954 \times 0.8 + 0.046 \times (1 - 0.9) = 0.768$$

- Our **forecast** of being in state 2 **next** period is

$$P(s_{t+1} = 2 | y_t) = 0.954 \times (1 - 0.8) + 0.046 \times 0.9 = 0.232$$

Bayesian updating with two states: examples IV

- Similarly, the mean and variance forecasts in this case are given by

$$
\begin{aligned}
E[y_{t+1}|y_t] &= \mu_1 P(s_{t+1} = 1|y_t) + \mu_2 P(s_{t+1} = 2|y_t) \\
&= -3 \times 0.768 + 1 \times 0.232 = -2.07 \\
Var(y_{t+1}|y_t) &= \sigma_1^2 P(s_{t+1} = 1|y_t) + \sigma_2^2 P(s_{t+1} = 2|y_t) \\
&\quad + P(s_{t+1} = 1|y_t) \times P(s_{t+1} = 2|y_t)(\mu_2 - \mu_1)^2 \\
&= 5^2 \times 0.768 + 2^2 \times 0.232 + 0.768 \times 0.232 \times (1+3)^2 \\
&= 22.98
\end{aligned}
$$

Bayesian updating with two states: example (cont.)

- Suppose instead we observe a value $y_t = +1$. Then

$$
\begin{aligned}
p(y_t|s_t &= 1) = Normpdf(4/5) = 0.0579 \\
p(y_t|s_t &= 2) = Normpdf(0) = 0.1995 \\
P(s_t &= 1|y_t) = \frac{0.0579 \times 0.70}{0.0579 \times 0.70 + 0.1995 \times 0.30} = 0.4038 \\
P(s_t &= 2|y_t) = \frac{0.1995 \times 0.30}{0.0579 \times 0.70 + 0.1995 \times 0.30} = 0.5962
\end{aligned}
$$

- Now, we **reduce** the probability of being in state 1 from 70% to 40%, while we increase the chance of being in state 2 from 30% to 60%
- Our **forecasts** of being in states 1 and 2 **next** period are

$$
\begin{aligned}
P(s_{t+1} &= 1|y_t) = 0.4038 \times 0.8 + 0.5962 \times (1 - 0.9) = 0.3827 \\
P(s_{t+1} &= 1|y_t) = 0.4038 \times (1 - 0.8) + 0.5962 \times 0.9 = 0.6173
\end{aligned}
$$

Estimation of Markov switching models

- The MS model is neither Gaussian, nor linear: the state s_t might lead to changes in regression coefficients and the covariance matrix
- Two common estimation methods:
  - Maximum likelihood estimation (MLE)
  - Bayesian estimation using Gibbs sampler
- Filtered states: $P(s_t = i|\mathcal{I}_t)$: probability of being in state i at time t given information at time t, $\mathcal{I}_t$
- Smoothed states: $P(s_t = i|\mathcal{I}_T)$: probability of being in state i at time t given information at the end of the sample, $\mathcal{I}_T$
- Choice of number of states can be tricky. We can use AIC or BIC

Filtered states (Ang-Timmermann, 2012)

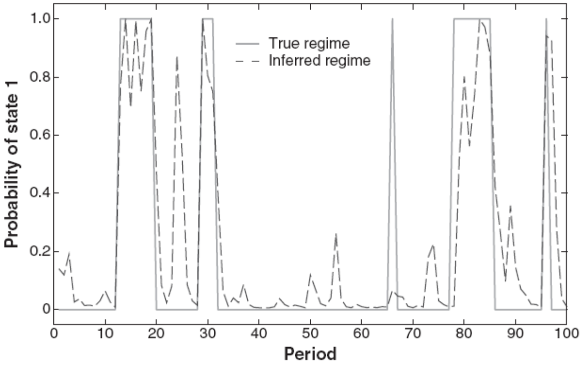
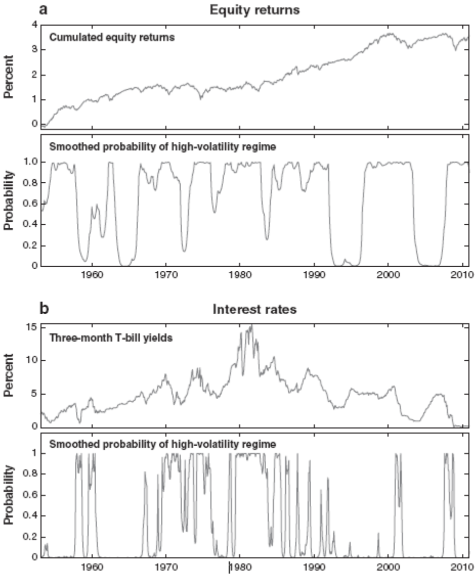


Figure 3

Mixture of normals. The figure plots a simulation from a regime-switching process with two states $s_t = 0$ and $s_t = 1$, with distributions $N(1, 1^2)$ and $N(-2, 2^2)$, respectively. The transition probabilities are $p_{00} = 0.95$ and $p_{11} = 0.80$. The true regime path is shown as the solid gray line, and the inferred (filtered) regime probability is graphed as the dashed darker gray line.

Smoothed state probabilities (Ang-Timmermann, 2012)



Smoothed state probabilities (Ang-Timmermann)

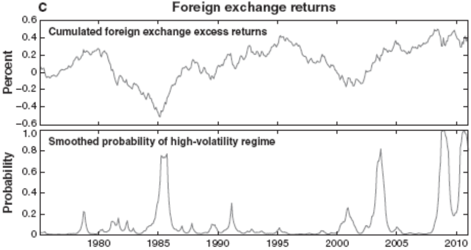


Figure 4

Smoothed probabilities. (*a*) Equity returns. (*b*) Interest rates. (*c*) Foreign exchange returns. In the bottom of each panel, we plot smoothed probabilities of being in regime $s_t = 0, p(s_t = 0 | I_T)$, conditional over the full sample computed following Hamilton (1990) and Kim (1994) from the regime-switching model (Equation 1) applied to equity excess returns, which are total returns (dividend plus capital gain) on the S&P500 in excess of the T-bills in panel *a*; interest rates, which are three-month T-bill yields in panel *b*; and foreign exchange excess returns (FX returns), which are returns from converting one USD into deutsch marks or euros, earning the German T-bill return, and then converting back to USD, in excess of the US T-bill return in panel *c*. The top of each panel shows cumulated sums of equity and FX returns in panels *a* and *c* and the three-month T-bill yield in panel *b*. All returns are at the monthly frequency. The sample period is 1953:01 to 2010:12 for equities and interest rates and 1975:01 to 2010:12 for foreign exchange returns.

Parameter estimates (Ang-Timmermann, 2012)

$$y_t = \mu_{s_t} + \phi_{s_t} y_{t-1} + \sigma_{s_t} \varepsilon_t, \quad \varepsilon_t \sim iiN(0,1)$$

Table 1 Parameter estimates[a]

	Equity returns		Interest rates		FX returns	
	Estimate	Std. error	Estimate	Std. error	Estimate	Std. error
μ_0	0.3326	0.2354	−0.0170	0.1273	0.4592	0.8921
μ_1	0.8994	0.2614	0.0355	0.0094	0.0143	0.0766
ϕ_0	0.0633	0.0460	0.9888	0.0018	−0.0392	0.1718
ϕ_1	−0.0426	0.0928	1.0000	0.0000	0.3033	0.0447
σ_0	4.8867	0.3448	0.8127	0.0512	4.3035	0.6314
σ_1	2.4462	0.4637	0.1760	0.0104	2.4842	0.1293
P	0.9770	0.0196	0.8789	0.0370	0.8692	0.1011
Q	0.9512	0.0206	0.9499	0.0147	0.9805	0.0179
	p-value		p-value		p-value	
Test $\mu_0 = \mu_1$	0.1057		0.7483		0.6203	
Test $\phi_0 = \phi_1$	0.0406		0.0000		0.0001	
Test $\sigma_0 = \sigma_1$	0.0160		0.0000		0.0026	

[a]We report parameter estimates of the regime-switching model (Equation 1) applied to equity excess returns, which are total returns (dividend plus capital gain) on the S&P500 in excess of the T-bills; interest rates, which are three-month T-bill yields; and foreign exchange excess returns (FX returns), which are returns from converting one USD into deutsch marks or euros, earning the German T-bill return, and then converting back to USD, in excess of the US T-bill return. All returns are at the monthly frequency. Estimations are done by maximum likelihood. The sample period is 1953:01 to 2010:12 for equities and interest rates and 1975:01 to 2010:12 for foreign exchange returns.

Take-away for MS models

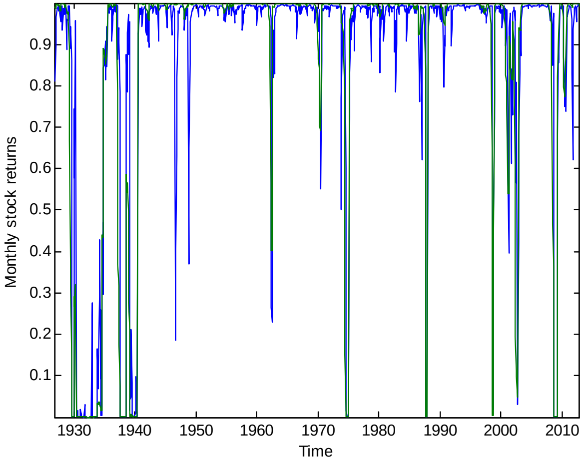
- Markov switching models are popular in finance and economics
- MS models are easy to interpret economically
  - Empirically often one state is highly persistent ("normal" state) with parameters not too far from the average of the series
  - The other state is often more transitory and captures spells of high volatility (asset returns) or negative outliers (GDP growth)
- Forecasts are easy to compute with MS models
- One state often has high volatility - regime switching can be important for risk management
- Try to experiment with the Markov switching and Kalman filter codes on Triton Ed

Estimates, 3-state model for monthly stock returns

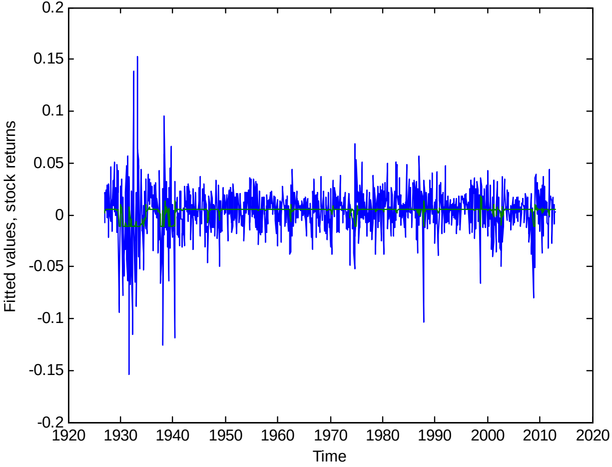
$$
\begin{aligned}
P' &= \begin{pmatrix} 0.9881 & 0.0119 & 0.000 \\ 0.000 & 0.9197 & 0.0803 \\ 0.8437 & 0.000 & 0.1563 \end{pmatrix} \\
\mu &= \begin{pmatrix} 0.0651 & -0.1321 & 0.3756 \end{pmatrix} \\
\sigma &= \begin{pmatrix} 0.0571 & 0.1697 & 0.0154 \end{pmatrix}
\end{aligned}
$$

- P : state transition probabilities, μ : means, σ : volatilities
- State 1: highly persistent, medium mean, medium volatility
- State 2: negative mean, high volatility, medium persistence
- State 3: transitory bounce-back state with high mean

Smoothed state probabilities, monthly stock returns



Fitted versus actual stock returns (3 state model)



Volatility of monthly stock returns