Introduction to Bayesian Methods

Julia Soloveva

Moscow State University

yulisoloveva-yu@yandex.ru

October 31, 2017

Overview

Let's meet

- Julia Soloveva
- Senior undergraduate student at MSU
- Fields of interest: Insurance Markets, Public Economics, Econometrics
- All questions are welcome throughout the lecture

Credits

Today's lecture is based on materials of the "Introductory Bayesian Time Series Methods" course at Barcelona Graduate School of Economics

Professor Gary Koop
University of Strathclyde

- Bayesian Basics
- Bayesian Methods for Fat Data
- Bayesian State Space Models

All relevant literature is listed later

Bayesian Philosophy - Blessed Uncertainty

Bayesian

1. Unknown things (parameters, models, forecasts) are random variables
2. We want distributions and means of unknown parameters
3. Few rules of probability

Frequentist

1. "True" values exist
2. We want point estimates and confidence intervals
3. Need to make assumptions (e.g. no unit roots)

Nothing is true, everything is permitted (c)

Bayesian Probability Rule

Definition (Conditional Probability)

The conditional probability of A given B, denoted as $Pr(A|B)$, is the probability of event A occurring given event B has occurred.

Conditional probability summarizes what is known about B given A
Well-known formulae:

$$Pr(A|B) = \frac{Pr(A,B)}{Pr(B)}$$

$$Pr(B|A) = \frac{Pr(A,B)}{Pr(A)}$$

Combined together will produce Bayes Theorem

Bayes Theorem

Theorem (Bayes Theorem)

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

Above is expressed in terms of two events, A and B. However, can be interpreted as holding true for random variables, A and B with probability density functions replacing Pr()s in previous formulae.

Econometrician's view

What does a humble researcher want?

- Estimate parameters in a model (regression coefficients)
- Compare different models (hypothesis testing)
- Predict (make forecasts)

What do we have being Bayes disciples?

- Let y be data, $y*$ be unobserved data (e.g. a forecast), M_i for $i = 1, .., m$ be a set of models, each of which depends on some parameters Θ^i.
- Model comparison based on posterior model probability: $p(M_i|y)$. Which model is the most likely?
- Prediction based on the predictive density $p(y|y^*)$.

Parameter Estimation in a Given Model

- Assume a single model which depends on parameters Θ $M(\Theta)$. We want to find properties (mean? distribution?) of the posterior $p(\Theta|y)$.
- Let us rewrite the Bayes rule:
  $p(B|A) = \frac{p(A|B)p(B)}{p(A)}$
  Replace B by Θ and A by y to obtain:
  $p(\Theta|y) = \frac{p(y|\Theta)p(\Theta)}{p(y)}$

Note: for estimation we can ignore $p(y)$ since it does not depend on Θ

- $p(\Theta|y)$ - the posterior. Given the data, what do we know about Θ?
- $p(y|\Theta)$ - likelihood function. How well do our parameters predict the data?
- $p(\Theta)$ - prior density. What do we know about parameters without seeing the data?

Prior Talk

Sounds unscientific?

- More information is good - if we do have prior information, we should include it.
- "Empirical Bayes" methods to estimate prior from the data
- Training sample prior - divide our sample into training sample and regression sample
- Prior sensitivity analysis

Prediction

Prediction is based on predictive density $p(y^*|y)$.

- Marginal density can be obtained from a joint density through integration:
  $p(y^*|y) = \int p(y^*, \Theta|y)d\Theta$
- Term inside integral can be rewritten as:
  $p(y^*|y) = \int p(y^*|y, \Theta)p(\Theta|y)d\Theta$
- Prediction involves the posterior and $p(y^*|y), \Theta)$

Model Comparison (Hypothesis Testing)

- Models denoted by M_i for $i = 1, .., m$. M_i depends on parameters Θ^i.
- $p(M_i|y)$ is *posterior model probability*
- Using Bayes rule with $B = M_i$ and $A = y$ we obtain:
  $p(M_i|y) = \frac{p(y|M_i)p(M_i)}{p(y)}$
- $p(M_i)$ - prior model probability
- $p(y|M_i)$ - marginal likelihood

Marginal Likelihood Calculation

Posterior can be written as: $p(\Theta^i|y, M_i) = \frac{p(y|\Theta^i, M_i)p(\Theta^i|M_i)}{p(y|M_i)}$
Integrate both sides with respect to Θ^i, use fact that
$\int p(\Theta^i|y, M_i)d\Theta^i = 1$ and rearrange: $p(y|M_i) = \int p(y|\Theta^i)p(\Theta^i|M_i)d\Theta^i$
Marginal likelihood depends only on the prior and the likelihood.

Model Comparison

For comparison of two models we calculate *Posterior odds ratio*

Definition (Posterior Odds Ratio)

$$PO_{ij} = \frac{p(M_i|y)}{p(M_j|y)} = \frac{p(y|M_i)p(M_i)}{p(y|M_i)p(M_j)}$$

Two Models Case

$$p(M_1|y) + p(M_2|y) = 1$$

$$PO_{12} = \frac{p(M_1|y)}{p(M_2|y)}$$

and we get: $p(M_1|y) = \frac{PO_{12}}{1+PO12}$

$$p(M_2|y) = 1 - p(M_1|y)$$

Definition (The Bayes Factor)

$BF_{ij} = \frac{p(y|M_i)}{p(y|M_j)}$

Why Bayes?

"Big" Data may be "tall" or "fat"

- Tall Data = data with many observations
- Fat Data = data with many variables

Fat Data is becoming common in macroeconomics.

Why include all the variables?

- The more data the better
- Want to avoid the omitted variable bias - misspecification
- Although the multicollinearity problem remains. How do we deal with it?

Bayes and Fat Data

Why not use conventional methods? Intuition:

- Number of observations, N, reflects amount of information in the data
- Number of variables, K reflects dimension of things trying to estimate with that data
- If N is small relative to K we're trying to do too much with too little information

What happens to least squares?

- Standard errors are growing, troubles with confidence intervals
- Fat data may be "noisy" - good in-sample fit, bad out-of-sample forecasting

Too Many Variables

Why don't we do hypothesis testing to reduce K? Because we get a *multiple testing problem*.

- With K=41 there are 41 different restricted regressions which drop one of the explanatory variables.
- There are $\frac{K(K-1)}{K}$ restricted regressions which drop two variables.
- etc.
- In total we have 2^K=2 199 023 255 552 possible regression models.

How do we choose? Obviously, no t-tests.

Bayesian Model Averaging - the Concept

We don't need no model selection. Away with single model and estimates based on it.

Model averaging - take a weighted average of estimates or forecasts from all models with weights given by $p(Mr|y)$.

- Let M_r for $r = 1, ..R$ denote R models
- If ϕ is a parameter to be estimated (or a function of parameters) or a variable to be forecast, then the rules of probability imply:

$$p(\phi|y) = \sum^R p(\phi|y, M_r)p(M_r|y)$$

- Allows for formal treatment of model uncertainty. Allows for uncertainty which model generated the data.

Dynamic Model Averaging

Another problem on the way: many time series variables.

Consider the problem: regression of inflation on many predictors. What may be the pitfalls?

- Regressions with many predictors can over-fit
- Marginal effects of predictors change over time (parameter change/structural breaks)
- The model itself may change
- Sometimes the random walk may be better!

Bayesians have an answer to that - Dynamic Model Averaging

The Final Example

- Consider a generalized Phillips curve: inflation depends on lagged inflation, unemployment and other predictors (change in consumption expenditures, change in real GDP, change in employment etc.).

One possible option - Time Varying Parameter Model (TVP)

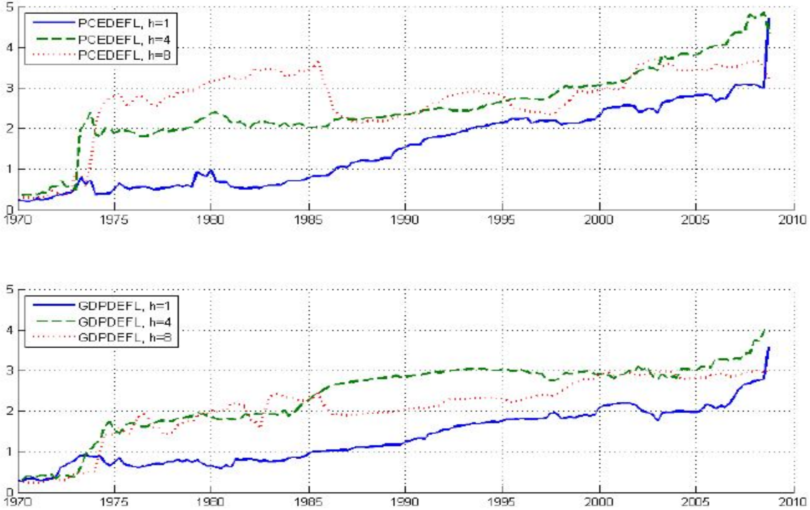
$y_t = z_t \Theta_t + \epsilon_t$
$\Theta_t = \Theta_{t-1} + \eta_t$
Where errors are i.i.d.

Disadvantage - same predictors are used at all points in time

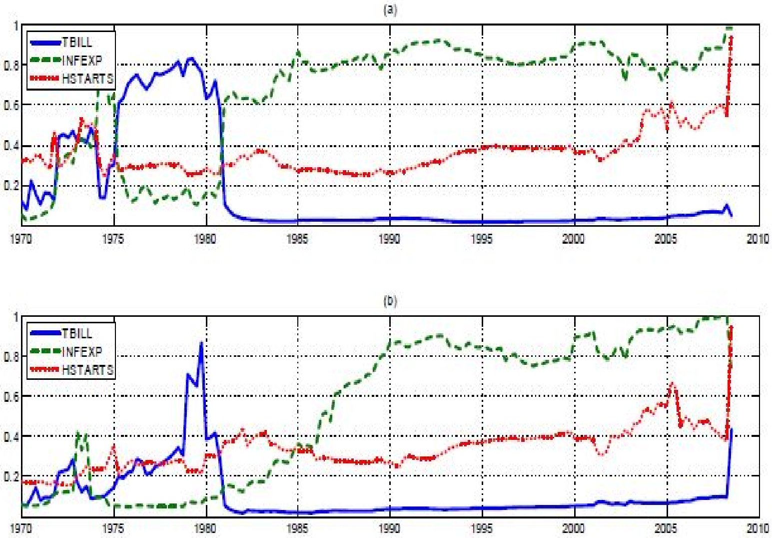
An example based on paper: Koop and Korobilis "Forecasting Inflation Using Dynamic Model Averaging" (2012, International Economic Review)

- Generalized Phillips Curve: inflation dependent on lagged inflation, unemployment and other predictors
- Other predictors: percentage changes in personal consumption, private residential fixed investment, real GDP; the log of housing starts, Treasury Bill market rate
- Two measures of inflation: core inflation, GDP deflator
- 3 forecasting horizons: $h = 1, 4, 8$

Dynamic Model Averaging - Results



Expected number of predictors



Posterior probability of inclusion of predictors

Textbooks:

- Gary Koop, Bayesian Econometrics
- Walter Enders, Applied Econometric Time Series; Barber, Cemgil, Chiappa Bayesian Time Series - for overview of Bayes theory in time series
- Giordani, Pitt, Kohn Bayesian Inference for Time Series State Space Models - an article on State Space Methods

Useful links:

- Course materials, tutorials and sample codes
- Bayesian Econometrics for Central Bankers

Boo!