Prior Predictive Analysis and Model Evaluation

Chapter 3, Complete and Incomplete Econometric Models

John Geweke
University of Technology Sydney and University of Colorado

SIRE Econometrics Workshop
University of Strathclyde
March 2, 2010

Outline

1. Prior predictive analysis
2. Example: Models of asset returns
3. Model evaluation
4. Example: Models of asset returns

Three components of a complete model A

- Prior density

$$p\left(\boldsymbol{\theta}_{A,T} \mid A\right)$$

- Conditional density of observables

$$p\left(\mathbf{y}_T \mid \boldsymbol{\theta}_{A,T}, A\right);$$

  - $L\left(\boldsymbol{\theta}_A; \mathbf{y}_T^o\right) = p\left(\mathbf{y}_T^o \mid \boldsymbol{\theta}_{A,T}, A\right)$ is the likelihood function.

- Vector of interest density

$$p\left(\boldsymbol{\omega} \mid \boldsymbol{\theta}_{A,T}, \mathbf{A}\right)$$

Simulating from the prior predictive distribution

- A vector of features of the observables

$$\begin{aligned} \mathbf{z}_T &= h(\mathbf{y}_T) \quad \text{(random variable)} \\ \mathbf{z}_T^o &= h(\mathbf{y}_T^o) \quad \text{(observed value)} \end{aligned}$$

- Predictive density is

$$p(\mathbf{z}_T|A) = \int p(\boldsymbol{\theta}_A)\, p(\mathbf{z}\,|A)\,.$$

- Simulation

$$\begin{aligned} \boldsymbol{\theta}_{A,T}^{(m)} &\sim p(\boldsymbol{\theta}_{A,T} \mid A) \\ \mathbf{y}_T^{(m)} &\sim p\left(\mathbf{y}_T \mid \boldsymbol{\theta}_{A,T}^{(m)}, A\right) \\ \mathbf{z}_T^{(m)} &= h\left(\mathbf{y}_T^{(m)}\right) \end{aligned}$$

Prior predictive analysis

- Recall the function

$$\mathbf{z}_T = h(\mathbf{y}_T)$$

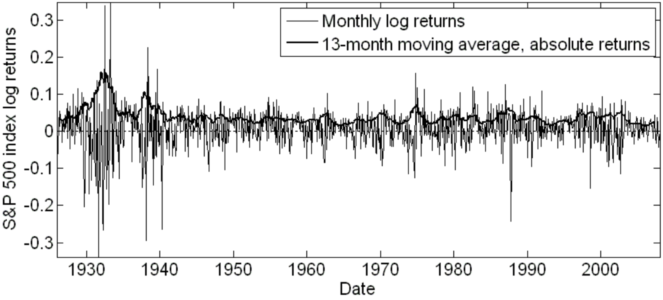
- Compare $p(\mathbf{z}_T | A)$, represented by

$$\mathbf{z}_T^{(m)} \sim p(\mathbf{z}_T \mid A) \quad (m = 1, \ldots, M)$$

  - with

$$\mathbf{z}_T^o = h(\mathbf{y}_T^o)$$

Data

- S&P 500 log returns
- 1926:1 - 2007:12
- 

Some observed features

Feature	Data
Return mean $\times$ 100	0.810
Return standard deviation $\times$ 100	5.515
Months in bear markets	350
Largest bear market decline	0.846
Return skewness	-0.437
Return excess kurtosis	8.166
Ratio of range to standard deviation	12.451
Return autocorrelation, lag 1	0.078
Squared return autocorrelation, lag 1	0.242
Squared return autocorrelation, lag 12	0.182
Absolute return long memory parameter	0.693

Some alternative models

- Let y_t denote the monthly return
- The whipping boy:

$$y_t \stackrel{iid}{\sim} N\left(\mu, \sigma^2\right)$$

- Mr. popularity:

$$y_t \sim N\left(\mu, \sigma_t^2\right); \;\; \sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

- The patriarch:

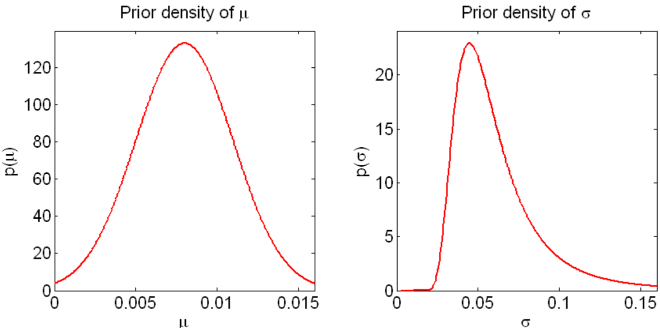
$$y_t \sim N\left(\mu, \sigma_t^2\right), \quad \log \sigma_t^2 = \alpha + \delta \log \sigma_{t-1}^2 + \sigma_v v_t, \;\; v_t \stackrel{iid}{\sim} N(0,1)$$

Prior distributions: Mean and variance

- Mean and variance in all three models:

$$\mu \sim N\left(.008, .003^2\right)$$
$$\frac{0.01}{\sigma^2} \sim \chi^2(4) \iff \frac{1}{\sigma^2} \sim \text{Gamma}(2, 200)$$

- The densities:



Prior distribution: Excess kurtosis
(GARCH and stochastic volatility models)

- Excess kurtosis:

  κ exponentially distributed with mean 8

- GARCH model:

$$y_t \sim N\left(\mu, \sigma_t^2\right); \ \ \sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

$$\kappa = \frac{6\alpha_1^2}{1 - \beta_1^2 - 2\alpha_1\beta_1 - 3\alpha_1^2} \geq 0$$

- Stochastic volatility model:

$$y_t \sim N\left(\mu, \sigma_t^2\right), \ \ \log \sigma_t^2 = \alpha + \delta \log \sigma_{t-1}^2 + \sigma_v v_t, \ \ v_t \overset{iid}{\sim} N(0,1)$$

$$\kappa = 3 \exp\left(\frac{\sigma_v^2}{1-\delta^2}\right) - 3 \geq 0$$

Prior distribution: Volatility persistence
(GARCH and stochastic volatility models)

- Correlation of squared return:

$$\rho \text{ uniformly distributed on } (0,1)$$

- GARCH model:

$$y_t \sim N\left(\mu, \sigma_t^2\right); \quad \sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

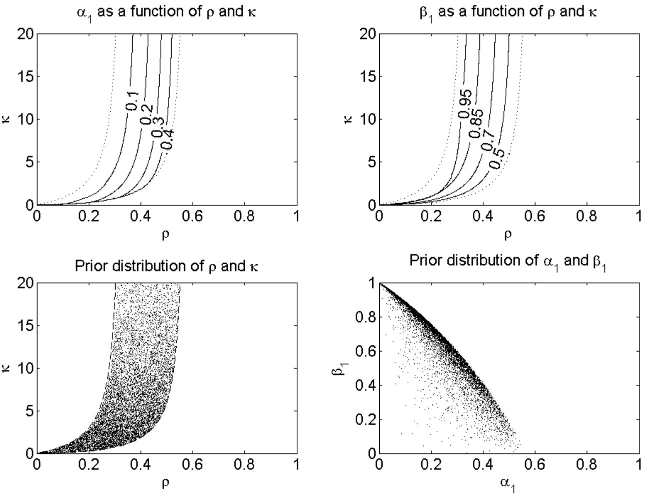
$$\rho = \frac{\alpha_1 \left(1 - \beta_1^2 - \alpha_1 \beta_1\right)}{1 - \beta_1^2 - 2\alpha_1 \beta_1}$$

- Stochastic volatility model:

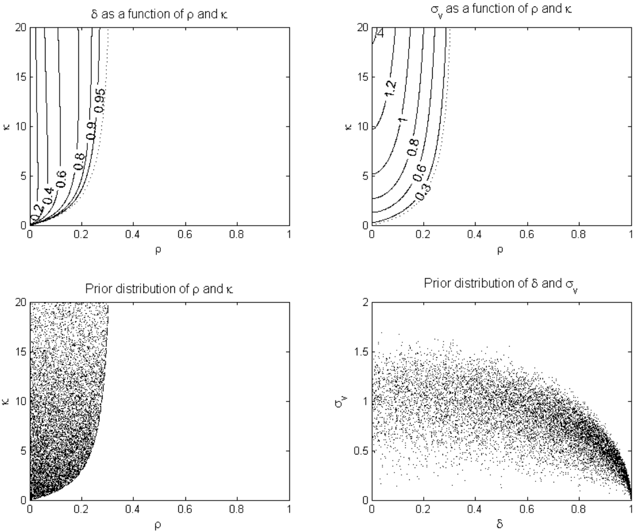
$$y_t \sim N\left(\mu, \sigma_t^2\right), \quad \log \sigma_t^2 = \alpha + \delta \log \sigma_{t-1}^2 + \sigma_v v_t, \quad v_t \overset{iid}{\sim} N(0,1)$$

$$\rho = \left[\exp\left(\frac{\delta \sigma_v^2}{1-\delta^2}\right) - 1\right] / \left[3 \exp\left(\frac{\sigma_v^2}{1-\delta^2}\right) - 1\right]$$

Priors in the GARCH model

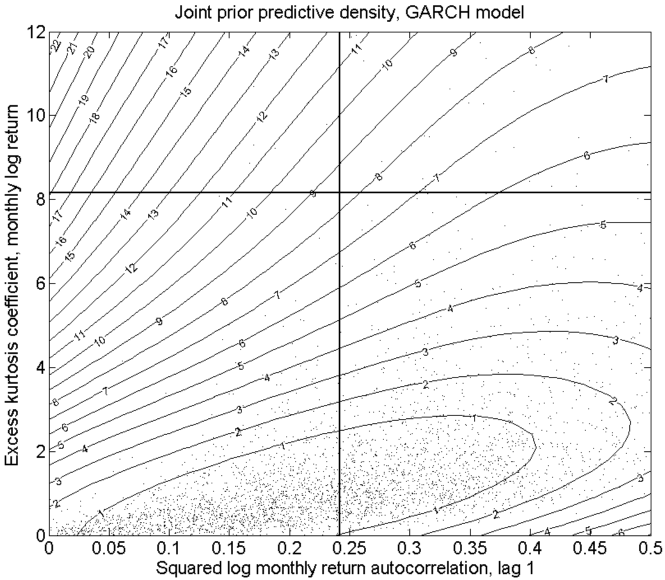


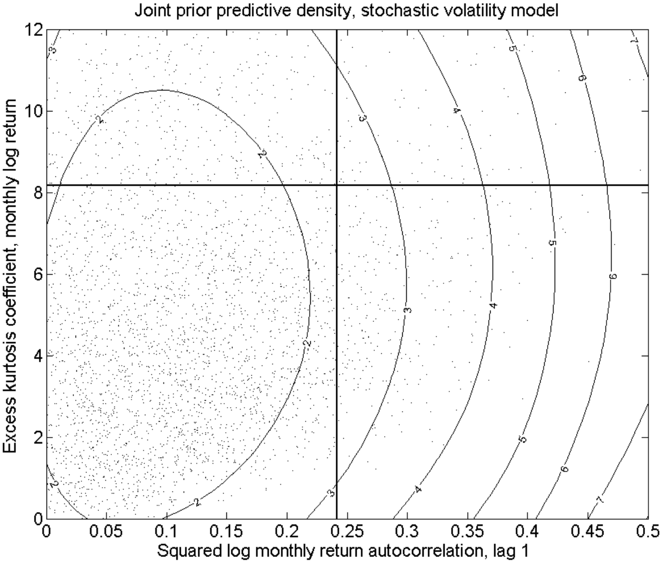
Priors in the stochastic volatility model



Prior predictive analysis: Overview

	Inverse cdf at data		
Feature	Gauss	GARCH	SV
Return mean	0.50	0.50	0.52
Return standard deviation	0.52	0.54	0.51
Months in bear markets	0.49	0.55	0.53
Largest bear market decline	0.91	0.90	0.91
Return skewness	0.00	0.03	0.16
Return excess kurtosis	1.00	0.98	0.74
Ratio of range to stan. dev.	1.00	0.66	0.68
Return autocorrelation, lag 1	0.99	0.96	0.96
Squared return AC, lag 1	1.00	0.51	0.88
Squared return AC, lag 12	1.00	0.91	0.98
$\lvert y_t \rvert$ long memory parameter	1.00	0.97	0.99


Joint prior predictive density, GARCH model
Excess kurtosis coefficient, monthly log return
Squared log monthly return autocorrelation, lag 1


Joint prior predictive density, stochastic volatility model
Excess kurtosis coefficient, monthly log return
Squared log monthly return autocorrelation, lag 1
0
2
4
6
8
10
12
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
0.45
0.5

A Bayesian alternative to pure significance testing:
Comparison with an incomplete model

The formalities

- A is a complete model under evaluation
- It therefore has a prior predictive distribution for all observables:
  - $\mathbf{y}$, all the data
  - $\mathbf{z} = g(\mathbf{y})$, a vector of features
  - In general, the prior distribution is $p(\mathbf{z} \mid A)$.
- Compare A with an incomplete model B
  - B does *not* provide $p(\mathbf{y} \mid B)$.
  - B *does* provide $p(\mathbf{z} \mid B)$

A Bayesian alternative to pure significance testing: Comparison with an incomplete model

The formalities (continued)

- What is the incomplete model B?
- Model B specifies a prior distribution for $\mathbf{z}$, but not for $\mathbf{y}$
  - Convenient
  - Not necessarily coherent
  - Mimics the way we think about what models ought to do
- See Geweke, *American Economic Review*, May 2007; *Complete and Incomplete Econometric Models*, Section 3.3.
- Examples to come.

A Bayesian alternative to pure significance testing:
Comparison with an incomplete model

The formalities (continued)

- Model A provides $p\left(\mathbf{z}^{o}|A\right)$.
  - Some care is required in computing this value
  - Matlab software available
- Model B also provides $p\left(\mathbf{z}^{o}|B\right)$.
- The Bayes factor in favor of model A is $p\left(\mathbf{z}^{o}|A\right)/p\left(\mathbf{z}^{o}|B\right)$.
  - Bayes factor large: Model A worth entertaining – write software
  - Bayes factor small: Model A deficient – keep working on model specification

The incomplete model

	Incomplete model prior distribution		
Feature	Functional form	Mean	St. dev.
Return mean	Gaussian	0.8	0.3
Return stan. dev.	σ^{-2} ~Gamma	0.62	0.31
Months in bear markets	Uniform(0, T)	$T/2$	$12^{-1/2}$T
Max. bear market decline	$\propto (1-f)^2 I_{(0.2,1)}(f)$	0.4	0.155
Return skewness	Gaussian	0	1
Return excess kurtosis	Exponential	8	8
Ratio of range to stan. dev.	1 + Gamma;	7	3
Return AC, lag 1	Gaussian	0	0.1
Squared return AC, lag 1	Uniform $(0,1)$	1/5	$12^{1/2}$
Squared return AC, lag 12	Beta	0.25	0.25
$\|y_t\|$ long memory parameter	Gaussian	0.5	0.25

Log Bayes factors in favor of complete model A over the incomplete model B

	Gaussian	GARCH	SV
Price drift (4 features)	-2.23	1.48	0.83
	-2.27	1.28	1.10
Return moments (3 features)	$-\infty$	-2.65	3.52
	$-\infty$	-2.14	3.48
Return dynamics (4 features)	$-\infty$	-0.04	2.46
	$-\infty$	-0.29	2.81
All 11 features	$-\infty$	1.25	4.44
	$-\infty$	1.15	3.82