



University of Zurich

Socioeconomic Institute
Sozialökonomisches Institut

Working Paper No. 0719

**Why Bayes Rules: A Note on Bayesian vs. Classical
Inference in Regime Switching Models**

Dennis Gärtner

December 2007

Socioeconomic Institute
University of Zurich

Working Paper No. 0719

Why Bayes Rules: A Note on Bayesian vs. Classical Inference in Regime Switching Models

December 2007

Author's address: Dennis Gärtner
E-mail: dennis.gaertner@soi.uzh.ch

Publisher Sozialökonomisches Institut
Bibliothek (Working Paper)
Rämistrasse 71
CH-8006 Zürich
Phone: +41-44-634 21 37
Fax: +41-44-634 49 82
URL: www.soi.uzh.ch
E-mail: soilib@soi.uzh.ch

Why Bayes Rules: A Note on Bayesian vs. Classical Inference in Regime Switching Models*

Dennis Gärtner[†]
University of Zürich

December 2007

Abstract

By means of a very simple example, this note illustrates the appeal of using Bayesian rather than classical methods to produce inference on hidden states in models of Markovian regime switching.

JEL Classification: C11, C22.

Keywords: Bayesian analysis, switching regression, regime changes, nonlinear filtering.

*The idea for this paper arose from joint work with Daniel Halbheer (see Gärtner and Halbheer, 2005). I am grateful to James Hamilton for very helpful comments.

[†]University of Zürich, Socioeconomic Institute, Blümlisalpstr. 10, CH-8006 Zurich, Switzerland. email: dennis.gaertner@soi.uzh.ch.

1 Introduction

Since Hamilton's (1989) seminal paper, models of Markovian regime switching have been widely applied in modeling all sorts of data. The classical estimation method originally proposed by Hamilton involves a two-step procedure in which model parameters are estimated first (usually by maximum likelihood), and inference on hidden states is subsequently drawn holding these parameter estimates fixed.

Advances in computational capacity have more recently spurred a number of papers employing alternative, Bayesian estimation methods based on Monte-Carlo techniques.¹ In contrast to classical methods, these methods permit simultaneous inference on both the model parameters and hidden states.

To our knowledge, despite the rising popularity of these newer techniques, little attempt has been made thus far to explicitly pinpoint their advantage over classical methods to an applied audience. In this note, we provide a very simple example to demonstrate the intuitive appeal of using the Bayesian approach.

2 A Brief Sketch of the Abstract Issue

The general problem considered has the following structure: We are concerned with a series of observations $\mathcal{Y}_T = (y_1, \dots, y_T)$, drawn from a distribution $p(\mathcal{Y}_T | \boldsymbol{\theta}, \mathcal{S}_T)$, where $\boldsymbol{\theta}$ denotes a vector of unknown model parameters, and $\mathcal{S}_T = (S_1, \dots, S_T)$ denotes a sequence of unobserved states. Ultimately, given a realization of \mathcal{Y}_T , we are interested in inferring $\boldsymbol{\theta}$ and \mathcal{S}_T .

The *classical* approach to this involves a two-step procedure: In a first step, we obtain a point estimate $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ (typically the maximum-likelihood estimate). Then, in a second step, we conduct inference on the state sequence given the data *and the parameter estimate* by calculating $p(\mathcal{S}_T | \mathcal{Y}_T, \hat{\boldsymbol{\theta}})$.

The *Bayesian* approach, on the other hand, treats $\boldsymbol{\theta}$ and \mathcal{S}_T as random variables and relies on calculating their posterior distribution $p(\mathcal{S}_T, \boldsymbol{\theta} | \mathcal{Y}_T)$. State inference is subsequently drawn based on the marginal posterior distribution of states, $p(\mathcal{S}_T | \mathcal{Y}_T)$.

¹See Kim and Nelson (1999b) for an introduction to these techniques; recent applications of these techniques include, for instance, Kim and Nelson (1999a), Kim and Nelson (2001), Gärtner and Halbheer (2005), Smith and Summers (2005).

In a Bayesian framework, the measures resulting from the two approaches can be related by

$$p(\mathcal{S}_T|\mathcal{Y}_T) = \int_{\boldsymbol{\theta}} p(\mathcal{S}_T|\mathcal{Y}_T, \boldsymbol{\theta})p(\boldsymbol{\theta}|\mathcal{Y}_T)d\boldsymbol{\theta}, \quad (1)$$

where $p(\boldsymbol{\theta}|\mathcal{Y}_T)$ denotes the posterior distribution of $\boldsymbol{\theta}$. Thus, differences in the two measures will ensue whenever (i) the researcher deems parameter constellations other than the point estimate $\hat{\boldsymbol{\theta}}$ likely, and (ii) such alternative parameter constellations are associated with a different evolution of the state sequence.

Obviously, correctly interpreted, neither approach is ‘wrong’. However, the point to be made by the example below is that the classical approach is much less amenable to a meaningful inference on states because any such inference is *conditional on the parameter estimate $\boldsymbol{\theta}$* .

As a case in point, Markovian regime switching models have been used extensively to detect booms and recessions in macroeconomic series (cf. Kim and Nelson, 1999a). In such a context, one is eventually interested in knowing the state the economy was in at a certain point in time, or whether the economy moved from one state to another. Conditioning such inference on the parameter estimates is then economically meaningful only to the extent that the researcher is very confident of these estimates.

In the next section, we illustrate this point by means of a highly stylized example which permits an intuitive grasp of the difference between conditional and unconditional state inference. In Section 4, we highlight the *practical* relevance of this point by showing that, also in less contrived examples, the conditional and unconditional measures can differ substantially.

3 A Simple Example

Consider a series of three observations, $\mathcal{Y}_T = (y_1, y_2, y_3)$, with

$$y_1 = 10, \quad y_2 = 15, \quad \text{and} \quad y_3 = 20. \quad (2)$$

Assume that the observations represent independent draws from a mixture of two normal distributions: Each y_t is drawn independently from either $N(\mu_1, \sigma)$ or from $N(\mu_2, \sigma)$, where $\mu_2 > \mu_1$. Which of these two distributions each observation t is drawn from is determined by the unobserved state $S_t \in \{1, 2\}$, so $y_t \sim N(\mu_{S_t}, \sigma)$. We will assume that this state sequence $\mathcal{S}_T = (S_1, S_2, S_3)$ itself represents an independent draw

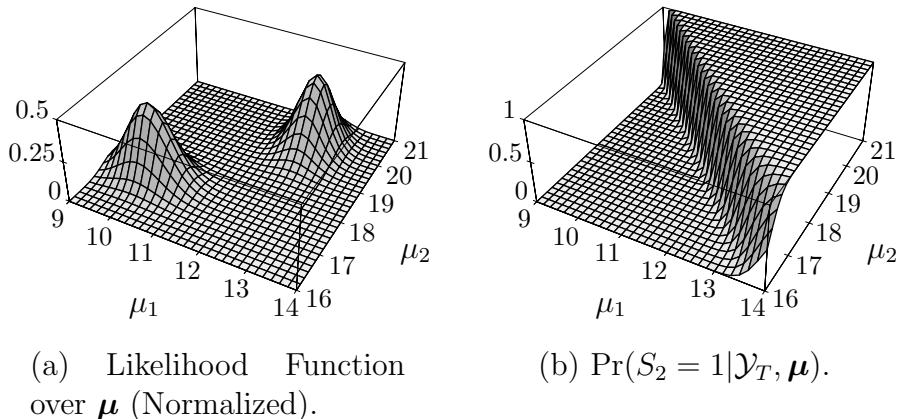


Figure 1: The Likelihood Function and Parameter-Contingent State Inference in the Example.

from $\{1, 2\}$, each with equal probability. Moreover, to keep things simple, we shall assume it known that $\sigma = 0.5$.

By simple intuition, there are essentially two combinations of parameters and states by which this model may have produced the observations in (2):

$$\textit{Scenario A: } \mathcal{S}_t = \{1, 2, 2\}, \mu_1 \approx 10, \mu_2 \approx 17.5;$$

$$\textit{Scenario B: } \mathcal{S}_t = \{1, 1, 2\}, \mu_1 \approx 12.5, \mu_2 \approx 20.$$

That is, it is rather clear that the lowest observation y_1 was drawn in state 1, whereas the highest observation y_3 was drawn in state 2.² What is unclear is how y_2 was produced: it may have been drawn from the same distribution as *either* y_1 or y_3 , leading to different ‘best guesses’ of μ_1 and μ_2 . Moreover, given that y_2 lies exactly half-way in between the other two observations, either scenario appears equally likely. Intuitively, therefore, inference should put equal probability on y_2 having been drawn from either distribution. That is, one would not expect the data to lead to any conclusion regarding the state in period two.

This intuition is confirmed by the likelihood function produced by the data over $\boldsymbol{\mu} = (\mu_1, \mu_2)$, which is shown in Figure 1(a) (normalized to integrate to 1). It displays two pronounced humps, one peaking at $(10, 17.5)$, the other at $(12.5, 20)$. Moreover, the parameter-contingent

²This insight makes use of the standard deviation σ being known and ‘rather low’. More precisely, the lower σ , the more likely it is that the data were produced by the described two parameter and state constellations rather than any other.

t	Bayesian	Classical
	$\Pr(S_t = 2 \mathcal{Y}_T)$	$\Pr(S_t = 2 \mathcal{Y}_T, \boldsymbol{\mu}^{\text{ML}})$
1	0.0	0.0
2	0.5	1.0
3	1.0	1.0

Table 1: State Inference in the Bayesian and in the Classical Setting.

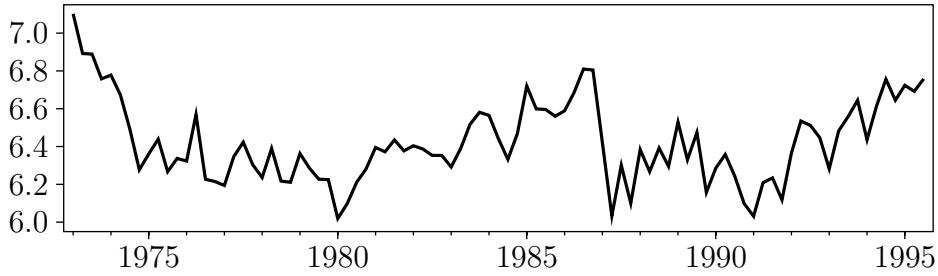
state probabilities depicted in panel (b) for $t = 2$ show that the first of these humps is associated with a very high likelihood of $S_2 = 1$, the second with a very high likelihood of $S_2 = 2$ (the corresponding plots of $\Pr(S_t = 1|\mathcal{Y}_T, \boldsymbol{\mu})$ for $t = 1$ and $t = 3$ are essentially level at 1.0 and 0.0, respectively, over the depicted parameter range).³ Thus, the humps in the likelihood function correspond to our two ‘scenarios’ above.

Next, let us see how classical and Bayesian estimation methods meet up with the above intuition. Under the Bayesian approach, the data \mathcal{Y}_T are combined with a prior $p(\boldsymbol{\mu}, \mathcal{S}_T)$ to produce the joint posterior distribution $p(\boldsymbol{\mu}, \mathcal{S}_T|\mathcal{Y}_T)$. Using an uninformative prior, the marginal posterior density for the parameters, $p(\boldsymbol{\mu}|\mathcal{Y}_T)$, is proportional to the likelihood surface shown in Figure 1. Moreover, Bayesian inference produces the marginal probabilities on states, $\Pr(S_t|\mathcal{Y}_T)$, shown in the center column of Table 1. Note that these figures are perfectly in line with our above intuition.⁴

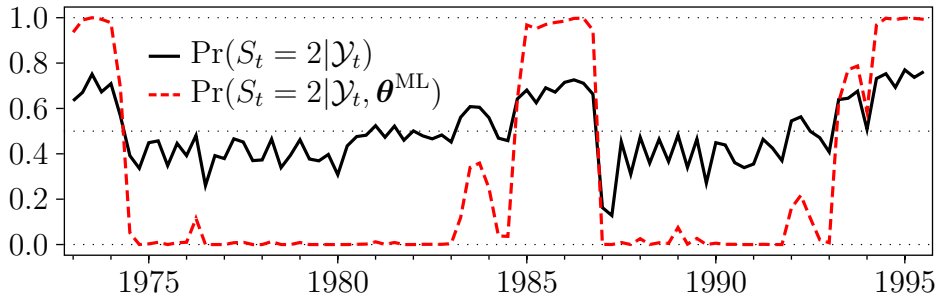
Next, consider the classical approach. As illustrated in Figure 1, the likelihood surface displays two peaks at equal height, either of which presents a valid maximum-likelihood parameter estimate. For specificity, let us use $\boldsymbol{\mu}^{\text{ML}} = (\mu_1^{\text{ML}}, \mu_2^{\text{ML}}) = (10, 17.5)$ as the ML-estimate. State inference in the classical setting is then based upon the probability of a certain state given both the data *and* $\boldsymbol{\mu}^{\text{ML}}$, $\Pr(S_t = 2|\mathcal{Y}_T, \boldsymbol{\mu}^{\text{ML}})$. These figures are reported in the rightmost column of Table 1. For $t = 2$, they differ markedly from both our above intuition and the figures obtained from the Bayesian analysis.

³Letting $f_{\text{N}(\mu, \sigma)}(x)$ denote the density function of a $\text{N}(\mu, \sigma)$ -distribution, the likelihood function in the example is simply $L(\boldsymbol{\mu}, \mathcal{Y}_T) = \prod_{t=1}^3 [\frac{1}{2} \cdot f_{\text{N}(\mu_1, \sigma)}(y_t) + \frac{1}{2} \cdot f_{\text{N}(\mu_2, \sigma)}(y_t)]$, whereas parameter-contingent state inference is obtained as $\Pr(S_t = i|\mathcal{Y}_T, \boldsymbol{\mu}) = f_{\text{N}(\mu_i, \sigma)}(y_t) / \sum_{j=1}^2 f_{\text{N}(\mu_j, \sigma)}(y_t)$.

⁴The derivation of $\Pr(S_t = 2|\mathcal{Y}_T)$ can be illustrated graphically in Figure 1: By equation (1), $\Pr(S_t = 2|\mathcal{Y}_T)$ is obtained by integrating up $\Pr(S_t = 2|\mathcal{Y}_T, \boldsymbol{\mu})$, shown in Figure 1(b), over the parameter space, with weights given by the (normalized) likelihood function shown in Figure 1(b).



(a) Log Merger Series



(b) Inferred Probability of Being in a High-Activity State

Figure 2: State Inference on US Log Merger Series 1973:I–1995:III under the Classical and the Bayesian Approach.

To appreciate this difference, note that the precise message of the conditional probability obtained from the classical analysis is the following: “Given that we believe μ^{ML} to be the true parameter values, the state in period 2 is almost certain to have been 2.” The example shows, however, that such a conditioning on parameter estimates can produce misleading conclusions regarding state inference: While the conditional probabilities obtained from the classical method suggest the clear identification of a state switch between observation 1 and 2, both intuition and the unconditional Bayesian estimates suggest that such a switch is equally likely *not* to have occurred.⁵

⁵In fact, the classical method would clearly *negate* a state-switch between the first two observations had we picked the other possible ML-parameter-estimate, (15, 17.5). The fundamental difficulty in interpreting the results from the classical method, however, is not immediately connected to the ambiguity in the ML-estimate: If observation y_2 were higher by an arbitrarily small amount, the ambiguity in the ML-estimate would disappear, whereas the interpretational caveat concerning conditional inference would obviously remain.

4 Conditional and Unconditional State Inference in Practice

To make the difference between the Bayesian and the classical approach particularly transparent, the above example was constructed so as to make it particularly easy to single out two relevant scenarios (i.e., likely combinations of states and parameter values). However, also in less contrived situations, the two measures of state inference above can differ substantially.

To illustrate this point, Figure 2 reports results from Gärtner and Halbheer (2005), where a (more elaborate) two-state Markov model is used to model the quarterly series of US mergers and acquisitions—the aim being the detection of periods of high merger activity (i.e., ‘merger waves’).

Results of both the Bayesian and the classical state inference are displayed in panel (b), showing a significant impact of conditioning state inference on the ML-parameter estimate θ^{ML} . Intuitively, the difference again stems from uncertainty concerning the parameters: Conditioning state inference on the point-estimate θ^{ML} neglects other likely parameter constellations which, apparently, are associated with alternate assessments of the state sequence. Given the richer data and parameter space, however, these alternative ‘scenarios’ (more technically: points of high density in the joint distribution of parameters and states) are of course no longer as easily identifiable.

5 Conclusion

By means of a simple example, we have illustrated the appeal of using Bayesian methods for inference on hidden states in models of Markovian state switching. By conditioning on a particular parameter estimate, state-probabilities obtained from classical methods can be misleading as regards inference on hidden states. The example illustrates that for these purposes, it is much more natural to employ the *unconditional* state probabilities which obtain from Bayesian methods.

References

- Gärtner, Dennis, and Daniel Halbheer (2005) ‘Are there waves in merger activity after all?’ Working Paper, University of Zürich
- Hamilton, James D. (1989) ‘A new approach to the economic analysis of nonstationary time series and the business cycle.’ *Econometrica* 57(2), 357–384
- Kim, Chang-Jin, and Charles R. Nelson (1999a) ‘Has the U.S. economy become more stable? a Bayesian approach based on a Markov-switching model of the business cycle.’ *Review of Economic Studies* 81(4), 608–616
- Kim, Chang-Jin, and Charles R. Nelson (1999b) *State-Space Models with Regime Switching* (Cambridge, Mass.: MIT Press)
- Kim, Chang-Jin, and Charles R. Nelson (2001) ‘A Bayesian approach to testing for Markov-switching in univariate and dynamic factor models.’ *International Economic Review* 42(4), 989–1013
- Smith, Penelope A., and Peter M. Summers (2005) ‘How well do Markov switching models describe actual business cycles? The case of synchronization.’ *Journal of Applied Econometrics* 20(2), 253–274

Working Papers of the Socioeconomic Institute at the University of Zurich

The Working Papers of the Socioeconomic Institute can be downloaded from http://www.soi.uzh.ch/research/wp_en.html

- 0719 Why Bayes Rules: A Note on Bayesian vs. Classical Inference in Regime Switching Models, December 2007, 8 p.
- 0718 Monoplistic Screening under Learning by Doing, Dennis Gärtner, December 2007, 29 p.
- 0717 An analysis of the Swiss vote on the use of genetically modified crops, Felix Schläpfer, November 2007, 23 p.
- 0716 The relation between competition and innovation – Why is it such a mess? Armin Schmutzler, November 2007, 26 p.
- 0715 Contingent Valuation: A New Perspective, Felix Schläpfer, November 2007, 32 p.
- 0714 Competition and Innovation: An Experimental Investigation, Dario Sacco, October 2007, 36p.
- 0713 Hedonic Adaptation to Living Standards and the Hidden Cost of Parental Income, Stefan Boes, Kevin Staub, Rainer Winkelmann, October 2007, 18p.
- 0712 Competitive Politics, Simplified Heuristics, and Preferences for Public Goods, Felix Schläpfer, Marcel Schmitt, Anna Roschewitz, September 2007, 40p.
- 0711 Self-Reinforcing Market Dominance, Daniel Halbheer, Ernst Fehr, Lorenz Goette, Armin Schmutzler, August 2007, 34p.
- 0710 The Role of Landscape Amenities in Regional Development: A Survey of Migration, Regional Economic and Hedonic Pricing Studies, Fabian Waltert, Felix Schläpfer, August 2007, 34p.
- 0709 Nonparametric Analysis of Treatment Effects in Ordered Response Models, Stefan Boes, July 2007, 42p.
- 0708 Rationality on the Rise: Why Relative Risk Aversion Increases with Stake Size, Helga Fehr-Duda, Adrian Bruhin, Thomas F. Epper, Renate Schubert, July 2007, 30p.
- 0707 I'm not fat, just too short for my weight – Family Child Care and Obesity in Germany, Philippe Mahler, May 2007, 27p.
- 0706 Does Globalization Create Superstars?, Hans Gersbach, Armin Schmutzler, April 2007, 23p.
- 0705 Risk and Rationality: Uncovering Heterogeneity in Probability Distortion, Adrian Bruhin, Helga Fehr-Duda, and Thomas F. Epper, July 2007, 29p.
- 0704 Count Data Models with Unobserved Heterogeneity: An Empirical Likelihood Approach, Stefan Boes, March 2007, 26p.
- 0703 Risk and Rationality: The Effect of Incidental Mood on Probability Weighting, Helga Fehr, Thomas Epper, Adrian Bruhin, Renate Schubert, February 2007, 27p.
- 0702 Happiness Functions with Preference Interdependence and Heterogeneity: The Case of Altruism within the Family, Adrian Bruhin, Rainer Winkelmann, February 2007, 20p.
- 0701 On the Geographic and Cultural Determinants of Bankruptcy, Stefan Buehler, Christian Kaiser, Franz Jaeger, June 2007, 35p.
- 0610 A Product-Market Theory of Industry-Specific Training, Hans Gersbach, Armin Schmutzler, November 2006, 28p.
- 0609 Entry in liberalized railway markets: The German experience, Rafael Lalive, Armin Schmutzler, April 2007, 20p.

- 0608 The Effects of Competition in Investment Games,
Dario Sacco, Armin Schmutzler, April 2007, 22p.
- 0607 Merger Negotiations and Ex-Post Regret,
Dennis Gärtner, Armin Schmutzler, September 2006, 28p.
- 0606 Foreign Direct Investment and R&D offshoring,
Hans Gersbach, Armin Schmutzler, June 2006, 34p.
- 0605 The Effect of Income on Positive and Negative Subjective Well-Being,
Stefan Boes, Rainer Winkelmann, May 2006, 23p.
- 0604 Correlated Risks: A Conflict of Interest Between Insurers and Consumers and Its
Resolution,
Patrick Eugster, Peter Zweifel, April 2006, 23p.
- 0603 The Apple Falls Increasingly Far: Parent-Child Correlation in Schooling and the
Growth of Post-Secondary Education in Switzerland,
Sandra Hanslin, Rainer Winkelmann, March 2006, 24p.
- 0602 Efficient Electricity Portfolios for Switzerland and the United States,
Boris Krey, Peter Zweifel, February 2006, 25p.
- 0601 Ain't no puzzle anymore: Comparative statics and experimental economics,
Armin Schmutzler, December 2006, 45p.
- 0514 Money Illusion Under Test,
Stefan Boes, Markus Lipp, Rainer Winkelmann, November 2005, 7p.
- 0513 Cost Sharing in Health Insurance: An Instrument for Risk Selection?
Karolin Becker, Peter Zweifel, November 2005, 45p.
- 0512 Single Motherhood and (Un)Equal Educational Opportunities: Evidence for Germany,
Philippe Mahler, Rainer Winkelmann, September 2005, 23p.
- 0511 Exploring the Effects of Competition for Railway Markets,
Rafael Lalive, Armin Schmutzler, April 2007, 33p.
- 0510 The Impact of Aging on Future Healthcare Expenditure;
Lukas Steinmann, Harry Telser, Peter Zweifel, December 2006, 23p.
- 0509 The Purpose and Limits of Social Health Insurance;
Peter Zweifel, September 2005, 28p.
- 0508 Switching Costs, Firm Size, and Market Structure;
Simon Loertscher, Yves Schneider, August 2005, 29p.
- 0507 Ordered Response Models;
Stefan Boes, Rainer Winkelmann, March 2005, 21p.
- 0506 Merge or Fail? The Determinants of Mergers and Bankruptcies in Switzerland, 1995-
2000; Stefan Buehler, Christian Kaiser, Franz Jaeger, March 2005, 18p.
- 0505 Consumer Resistance Against Regulation: The Case of Health Care
Peter Zweifel, Harry Telser, Stephan Vaterlaus, February 2005, 23p.
- 0504 A Structural Model of Demand for Apprentices
Samuel Mühlemann, Jürg Schweri, Rainer Winkelmann and Stefan C. Wolter,
February 2005, 25p.
- 0503 What can happiness research tell us about altruism? Evidence from the German
Socio-Economic Panel
Johannes Schwarze, Rainer Winkelmann, September 2005, 26p.
- 0502 Spatial Effects in Willingness-to-Pay: The Case of Two Nuclear Risks
Yves Schneider, Peter Zweifel, September 2007, 31p.
- 0501 On the Role of Access Charges Under Network Competition
Stefan Buehler, Armin Schmutzler, January 2005, 30p.
- 0416 Social Sanctions in Interethnic Relations: The Benefit of Punishing your Friends
Christian Stoff, February 2006, 51p.