Gibbs Sampling for a Duration Dependent Markov Switching Model with an Application to the U.S. Business Cycle

Matteo Pelagatti

1 Introduction and motivation

Does the probability of moving from a recession into an expansion depend on how long the economy has been in recession? Similarly, does the probability that the economy may fall into a recession depend on the length of the expansion phase? The present paper tries to answer these two questions, at least as far as the U.S. economy is concerned[1].

Some authors have already dealt with the duration-dependence problem: Diebold et al. [7] and Watson [14] apply nonparametric methods to the NBER's dating of business cycles, while Durland and McCurdy [8] use an extension of the Hamiton's [11] Markov switching model. Their conclusions are in favor of the duration dependence hypothesis, at least in the contraction phase. In particular Diebold et al. [7] conclude that, given their results, Hamilton's Markov switching model is miss-specified because its transition probabilities are constant over time; furthermore they also argue that including duration dependence in a Markov switching problem may raise identification and estimation problems.

In this paper we use a model in some way similar to the one used by Durland and McCurdy [8], but with some major differences: (i) we work within a Bayesian framework combined with MCMC (Markov Chain Monte Carlo) methods, (ii) we show that the duration-dependent switching model has a representation as time-invariant Markov switching model, allowing the use of all the standard tools of such models, (iii) we use a probit regression

[1] We chose to work with the U.S. GDP, as this time series has become a benchmark for new models for the business cycle analysis: the U.S. GDP has been thoroughly studied with many type of models, and when a new model, such as the one presented here, is proposed, it's thus easier to compare the results and draw conclusions.

instead of a logit regression to model the duration dependent probability of switching from one state to the other.

We chose to implement Bayesian MCMC methods since working with complex models with many unknown quantities and relatively few data in a maximum-likelihood (ML) framework has some drawbacks: (i) the whole ML inference is based on asymptotic results, which in models with so many parameters (relative to the amount of data) may be poor approximations, (ii) the likelihood surface can be rather flat and local minima problems may arise, (iii) the inference on latent variables (the state of the economy) is done (by means of filters) conditional on the estimated parameters, and therefore it does not reflect the parameters uncertainty. Problem (i) is significantly reduced in Bayesian MCMC methods, in fact the approximation to the posterior distribution of the parameters can virtually be made as good as wished. Problem (ii) has to do with the uncertainty that the scarce amount of data leaves when the model is complex; the use of prior information can be crucial to reduce such uncertainty. Problem (iii) does not arise in our Bayesian-MCMC framework, as the posterior distributions of every unknown quantity is simulated.

2 The model

The duration-dependent switching model that we will use is the following

$$\phi(L)(y_t - \mu_0 - \mu_1 S_t) = \varepsilon_t \qquad t = 1, \ldots, T \tag{1}$$

where

- $\phi(L) = 1 - \phi_1 L - \ldots - \phi_r L^r$ is a stationary autoregressive polynomial in the lag operator L ($L^j x_t = x_{t-j}$),
- $\{S_t\}$ is a 0-1 Markov process with transition matrix $\mathbf{P}$, and will be thoroughly defined in section 2.1 in a way that allows duration dependence of transition probabilities,
- $\{\varepsilon_t\}$ is a Gaussian white noise process with variance σ^2.

The unknown quantities of the model are $\phi_1, \ldots, \phi_r, \mu_0, \mu_1, \sigma^2, \mathbf{P}, \{S_t\}_1^T$.

2.1 States and durations

Let $\{0, 1\}$ be the state space of S_t, 0 being the contraction state and 1 the expansion state. In the usual Markov switching model S_t is a homogeneous

Markov chain with transition matrix $\mathbf{P}$. In order to include duration dependence in the model we can proceed in two different ways: (i) S_t can be modeled as a non-homogeneous Markov chain, where the transition matrix changes according to the duration of each state (i.e. $\mathbf{P}_t = \mathbf{P}(D_{t-1})$, with D_t state-duration variable), (ii) the pair (S_t, D_t), with D_t state-duration variable, can be modelled as a homogeneous Markov chain, and the variable of interest S_t can be recovered through marginalization with repect to D_t. Here we will use the latter solution, which has the advantage of preserving the theoretical framework of Markov switching models.

Let D_t be the duration variable, a variable that counts the number of times in which S_t remains in the same state (an example is given in table 1). We want to build a model where the probability of S_t being in a particular

t	1	2	3	4	5	6	7	8	9	10	11	12
S_t	1	1	1	1	0	0	0	1	0	0	0	0
D_t	3	4	5	6	1	2	3	1	1	2	3	4

Table 1: A possible realization of processes S_t and D_t.

state depends only on the previous state S_{t-1} and duration D_{t-1}; but given D_{t-1} and S_t, D_t is completely determined ($D_t = D_{t-1} + S_t$ if $S_t = S_{t-1}$, $D_t = 1$ if $S_t \neq S_{t-1}$) and therefore the pair (S_t, D_t) is a Markov chain with state space

$$\{(0,1),(1,1),(0,2),(1,2),(0,3),(1,3),\ldots\}$$

and transition matrix[2]

$$\mathbf{P} = \begin{bmatrix} 0 & p_{0|1}(1) & 0 & p_{0|1}(2) & 0 & p_{0|1}(3) & \ldots \\ p_{1|0}(1) & 0 & p_{1|0}(2) & 0 & p_{1|0}(3) & 0 & \ldots \\ p_{0|0}(1) & 0 & 0 & 0 & 0 & 0 & \ldots \\ 0 & p_{1|1}(1) & 0 & 0 & 0 & 0 & \ldots \\ 0 & 0 & p_{0|0}(2) & 0 & 0 & 0 & \ldots \\ 0 & 0 & 0 & p_{1|1}(2) & 0 & 0 & \ldots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix},$$

where $p_{i|j}(d) = \Pr(S_t = i | S_{t-1} = j, D_{t-1} = d)$.

In order to work with finite state space and transition matrix, we fix a finite maximum value for the support of D_t, say τ, and conditional on $D_{t-1} = \tau$ we assign non-zero probabilities only to the four events

$$\{(S_t = i, D_t = \tau) | (S_{t-1} = i, D_{t-1} = \tau)\},$$

[2] The transition matrix is here designed so that the rows, and not the columns, sum to one.

$$\{(S_t = i, D_t = 1)|(S_{t-1} = j, D_{t-1} = \tau)\},$$
$$i, j = \{0, 1\}, i \neq j.$$

This assumption implies that, when the economy has been in state i at least τ times, the additional times in which the true state remains i influence no more the probability of transition (i.e. $p_{j|i}(\tau) = p_{j|i}(\tau + n)$, with $i, j = 0, 1$ and n positive integer).

As pointed out by Hamilton ([12], section 22.4) we can always write the likelihood function of y_t depending only on the state variable at time t, even though in our model a r-order autoregressive component is present; this can be done using the state variable $S_t^* = (D_t, S_t, S_{t-1}, \ldots, S_{t-r})$ that comprehend also the possible combinations of the states of the economy in the last r periods. In table 2 the state space of S_t^* when $r = 4$ and $\tau = 5$ is shown. If $\tau \geq r$ the maximum number of non-negligible states[3] is given by $\sum_{i=1}^{r} 2^i + 2(\tau - r)$. The transition matrix $\mathbf{P}^*$ of such a state variable is sparse and quite straightforward to build.

2.2 Probit model for the probabilities of transition

The transition matrix $\mathbf{P}^*$ of the Markov chain S_t^* contains 2τ parameters to estimate. In order to give a more parsimonious parametrization, we use the following Probit model.

Consider the linear model

$$S_t^\bullet = [\beta_1 + \beta_2 D_{t-1}]S_{t-1} + [\beta_3 + \beta_4 D_{t-1}](1 - S_{t-1}) + \epsilon_t \tag{2}$$

with $\epsilon_t \sim \mathcal{N}(0, 1)$, and $S_t^\bullet$ latent variable defined by

$$\Pr(S_t^\bullet \geq 0|S_{t-1}, D_{t-1}) = \Pr(S_t = 1|S_{t-1}, D_{t-1}) \tag{3}$$
$$\Pr(S_t^\bullet < 0|S_{t-1}, D_{t-1}) = \Pr(S_t = 0|S_{t-1}, D_{t-1}). \tag{4}$$

Thus the following events are equivalent

$$\begin{aligned}
&\{S_t^\bullet \geq 0|S_{t-1}, D_{t-1}\} \equiv \\
&\equiv \{[\beta_1 + \beta_2 D_{t-1}]S_{t-1} + [\beta_3 + \beta_4 D_{t-1}](1 - S_{t-1}) + \epsilon_t \geq 0\} \equiv \\
&\equiv \{\epsilon_t \geq [-\beta_1 - \beta_2 D_{t-1}]S_{t-1} + [-\beta_3 - \beta_4 D_{t-1}](1 - S_{t-1})\}
\end{aligned}$$

and it holds

$$\begin{aligned}
p_{1|1}(d) &= \Pr(S_t = 1|S_{t-1} = 1, D_{t-1} = d) = \qquad (5)\\
&= 1 - \Phi(-\beta_1 - \beta_2 d) \\
p_{0|0}(d) &= \Pr(S_t = 0|S_{t-1} = 0, D_{t-1} = d) = \Phi(-\beta_3 - \beta_4 d) \qquad (6)
\end{aligned}$$

[3]'Negligible states' stands here for 'states always associated with zero probability'.

	D_t	S_t	S_{t-1}	S_{t-2}	S_{t-3}	S_{t-4}
1	1	0	1	0	0	0
2	1	0	1	0	0	1
3	1	0	1	0	1	0
4	1	0	1	0	1	1
5	1	0	1	1	0	0
6	1	0	1	1	0	1
7	1	0	1	1	1	0
8	1	0	1	1	1	1
9	1	1	0	0	0	0
10	1	1	0	0	0	1
11	1	1	0	0	1	0
12	1	1	0	0	1	1
13	1	1	0	1	0	0
14	1	1	0	1	0	1
15	1	1	0	1	1	0
16	1	1	0	1	1	1
17	2	0	0	1	0	0
18	2	0	0	1	0	1
19	2	0	0	1	1	0
20	2	0	0	1	1	1
21	2	1	1	0	0	0
22	2	1	1	0	0	1
23	2	1	1	0	1	0
24	2	1	1	0	1	1
25	3	0	0	0	1	0
26	3	0	0	0	1	1
27	3	1	1	1	0	0
28	3	1	1	1	0	1
29	4	0	0	0	0	1
30	4	1	1	1	1	0
31	5	0	0	0	0	0
32	5	1	1	1	1	1

Table 2: State space of $S_t^* = (D_t, S_t, S_{t-1}, \ldots, S_{t-p})$ for $p = 4$, $\tau = 5$.

where $d = 1, \ldots, \tau$ and $\Phi(.)$ is the standard normal cumulative distribution function. Now the four parameters $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3, \beta_4)'$ completely define the transition matrix $\mathbf{P}^*$, meaning that knowing $\boldsymbol{\beta}$, all the $p_{i|j}(d)$ (i,j={0,1}) are determined.

3 Gibbs sampling from the posterior distribution of the model's parameters

Gibbs sampling is a MCMC technique where the posterior distribution of the parameters is simulated by serially generating random quantities from their full conditional distributions (i.e. the posterior distribution of each parameter or vector of parameters given all the other parameters). In this section we show how to Gibbs sample from the posterior distribution of the parameters (including the latent variables) of our model.

Once we fix some initial values, the parameters' priors and the full conditional posterior distribution of the parameters, it's straightforward to Gibbs sample: it's a metter of generating random values from each conditional posterior distribution, given the last generated values of all the other parameters.

3.1 Prior distribution

In the next subsections we will show that with suitable transformations the problem of building the full conditional posterior distribution for our model's parameters in many cases reduces to the problem of finding the posterior of the parameters of a Gaussian linear model. In order to exploit conjugacy, we use (truncated) normal priors for the regression coefficients and inverse-gamma priors for the regression variance (see appendix).

The priors that will be used are

$$\begin{aligned}
\sigma_y^2 &\sim \mathcal{IG}(n_0, n_0 v_0) && (7)\\
\boldsymbol{\phi} &\sim \mathcal{N}(\mathbf{f}_0, \sigma_y^2 \mathbf{F}_0) I(\{\phi(L) \text{ stationary}\}) && (8)\\
\boldsymbol{\mu} &\sim \mathcal{N}(\mathbf{m}_0, \sigma_y^2 \mathbf{M}_0) I(\{\mu_0 \le \mu_1\}) && (9)\\
\boldsymbol{\beta} &\sim \mathcal{N}(\mathbf{b}_0, \mathbf{B}_0). && (10)
\end{aligned}$$

To complete the definition of the priors for our model, we also need the distribution of S_0^*, which can be done attributing suitable probabilities to each possible outcome; more details are given in section 3.2.3.

3.2 Gibbs sampling steps

Initial values for each parameter (ϕ's, μ's, σ^2, β's) and latent variable (s_t's) must be chosen. In the long run the initial values play no role, but good initial values can speed up the convergence of the Markov chain to it's ergodic distribution and avoid underflow problems to the computer's floating point processor.

The following steps are to be iterated until a sufficiently large sample is produced.

3.2.1 Step 1. ϕ's

Using the parameter values of last iteration (or initial values if this is the first iteration) calculate

$$\bar{y}_t = (y_t - \mu_0 - \mu_1 s_t) \tag{11}$$

Equation (1) can be rewritten as

$$\bar{y}_t = \phi_1 \bar{y}_{t-1} + \dots \phi_p \bar{y}_{t-p} + \varepsilon_t \tag{12}$$

which is a linear model. A value for the parameter vector $\boldsymbol{\phi}$ and a value for σ^2 are drawn from their posterior distribution obtained as shown in appendix[4].

3.2.2 Step 2. μ's

Using the newest values of the other parameters, calculate

$$\tilde{y}_t = \phi(L) y_t \tag{13}$$
$$c_t = \phi(1) \tag{14}$$
$$\tilde{s}_t = \phi(L) s_t. \tag{15}$$

Equation (1) can now be rewritten as

$$\tilde{y}_t = \mu_0 c_t + \mu_1 \tilde{s}_t + \varepsilon_t \tag{16}$$

which is a linear model with known σ_y^2 variance. A new value for the parameter vector $\boldsymbol{\mu}$ is drawn as shown in equations (28)–(30) of the appendix.

[4] The variance of $\tilde{y}_t$ needed in the formulae of the appendix is given by the first element of the matrix $\sigma^2[I_{r^2} - (\mathbf{F} \otimes \mathbf{F})]^{-1}$ with

$$\mathbf{F} = \begin{pmatrix} \phi_1 & \phi_2 & \dots & \phi_r \\ & & & 0 \\ & \mathbf{I}_{r-1} & & \vdots \\ & & & 0 \end{pmatrix}$$

3.2.3 Step 3. s_t's

We will use an implementation of the multi-move Gibbs sampler originally proposed by Carter and Kohn [4].

Let $\hat{\boldsymbol{\xi}}_{t|i}$ be the vector containing the probabilities of being in each state (the first element is the probability of being in state 1, the second element is the probability of being in state 2, and so on) given $\mathcal{Y}_i = (y_1, \dots, y_i)$ and the model's parameters.

Let $\boldsymbol{\eta}_t$ be the vector containing the likelihood of each state given $\mathcal{Y}_t$ and the model's parameters.

The Hamilton's ([12], section 22.4) filter recursion is given by

$$\hat{\boldsymbol{\xi}}_{t|t} = \frac{\hat{\boldsymbol{\xi}}_{t|t-1} \odot \boldsymbol{\eta}_t}{\hat{\boldsymbol{\xi}}'_{t|t-1} \boldsymbol{\eta}_t} \tag{17}$$

$$\hat{\boldsymbol{\xi}}_{t+1|t} = \mathbf{P}^* \hat{\boldsymbol{\xi}}_{t|t} \tag{18}$$

with the symbol $\odot$ indicating element by element multiplication. The filter is completed with the prior probabilities vector $\hat{\boldsymbol{\xi}}_{1|0}$, that, in absence of any better information can be set equal to the vector of ergodic (or stationary) probabilities of $\{S_t^*\}$.

In order to sample from the distribution of $\{S_t^*\}_1^T$ given the full information set $\mathcal{Y}_T$, it can be shown that

$$\Pr(S_t^* = i | \mathcal{Y}_T) = \Pr(S_t^* = i | S_{t+1}^* = k, \mathcal{Y}_t) = \frac{p_{i|k} \hat{\xi}_{t|t}^{(i)}}{\sum_{j=1}^m p_{j|k} \hat{\xi}_{t|t}^{(j)}}, \tag{19}$$

where $p_{j|k}$ indicates the transition probability of moving to state j from state k (element (j,k) in the transition matrix $\mathbf{P}^*$) and $\hat{\xi}_{t|t}^{(j)}$ is the j-th element of vector $\hat{\boldsymbol{\xi}}_{t|t}$. In matrix notation the same can be written as

$$\hat{\boldsymbol{\xi}}_{t|(S_{t+1}=k, \mathcal{Y}_T)} = \frac{\mathbf{P}^{*\,\prime}_{[k,.]} \odot \hat{\boldsymbol{\xi}}_{t|t}}{\mathbf{P}^*_{[k,.]} \hat{\boldsymbol{\xi}}_{t|t}} \tag{20}$$

where $\mathbf{P}^*_{[k,.]}$ identifies the k-th row of the transition matrix $\mathbf{P}^*$.

The states can now be generated starting from the filtered probability $\hat{\boldsymbol{\xi}}_{T|T}$ and proceeding backward $(T-1, \dots, 1)$, using equation (20) where k is to be substituted with the last generated value (s_{t+1}).

3.2.4 Step 4. β's

As illustrated by Albert and Chib [1] we can generate $\boldsymbol{\beta}$ using data augmentation. Given the generated s_t's and the four parameters β's of last iteration we can generate the $S_t^\bullet$'s of the Probit model of section 2.2 from the following truncated normals:

$$S_t^\bullet|(S_t = 0, \mathbf{x}_t, \boldsymbol{\beta}) \sim \mathcal{N}(\mathbf{x}_t'\boldsymbol{\beta}, 1) I_{(-\infty,0)} \tag{21}$$
$$S_t^\bullet|(S_t = 1, \mathbf{x}_t, \boldsymbol{\beta}) \sim \mathcal{N}(\mathbf{x}_t'\boldsymbol{\beta}, 1) I_{[0,\infty)} \tag{22}$$
$$\tag{23}$$

with

$$\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3, \beta_4)'$$
$$\mathbf{x}_t = (s_{t-1}, s_{t-1}, d_{t-1}, (1 - s_{t-1}), (1 - s_{t-1})d_{t-1})'$$

and $I_{(a,b)}$ indicator variable used to denote truncation.

With the generated $s_t^\bullet$'s the Probit regression equation (2) become just a linear model with known unit variance. New values for $\boldsymbol{\beta}$ are generated using its posterior distribution (see appendix).

4 Application to the U.S. Business Cycle

The model (with $r = 4$, $\tau = 20$) and the inferential methodology that have been presented in the previous sections, are now applied to the quarterly and seasonal adjusted U.S. GDP time series ranging from 1947Q1 to 2000Q3, as published by the Bureau of Economic Analysis. The $\{y_t\}$ in formulae is not the GDP time series itself but the following transformation: $y_t = 100 \cdot \Delta \ln(\mathrm{GDP}_t)$, Δ denoting differentiation. Such a transformation let y_t be an approximation of the quarterly percentage growth rate of the GDP and give the μ's of the model a straightforward interpretation as mean percentage growth rates: μ_0 is the contraction's mean growth rate, $\mu_0 + \mu_1$ is the expansion's mean growth rate.

4.1 Further selection of the model and prior distributions

Running the Gibbs sampler for the full model with rather vague priors and analyzing the marginal posterior distributions, the model does not seem to work too well: the 95%-credibility intervals of the posterior distributions of the ϕ's all comprehend the value 0 and the series of contraction probabilities does not clearly individuate contractions and expansions. With a deeper analysis of the bivariate distributions of each ϕ with each β we find out that there is some correlation (≈ 0.5) between them, and even a -0.95 correlation between β_3 and β_4. Some collinearity between the constant relative to the parameter β_3 and the duration variable relative to the parameter β_4 was expected, as recessions are often very short and therefore the duration variable does not move too much, but such a high correlation of the two parameters is not a good sign. We conclude that both the autoregressive part of the model and the duration-dependent Markov chain try to "explain" the same autocorrelation of the series, and therefore the likelihood surface result in some directions rather flat. To solve this problem, considering also the high concentration around the zero of the ϕ's posterior, we decide to eliminate the autoregressive part of the model, setting $\phi_1 = \ldots = \phi_4 = 0$. Albert and Chib [2], in an application of the Gibbs sampler to a standard autoregressive Markov switching model on a subset of our GDP time series, came to the same conclusion of excluding the autoregressive part of the model.

The prior distributions assigned to the vectors of parameters $\boldsymbol{\mu}$ and $\boldsymbol{\beta}$ are chosen not too tight, but precise enough to focus the sampling in an economically reasonable subset of the parameter space (see table 3). σ^2 is given a vague prior.

The Gibbs sampler is run for 10000 iterations, after a burn-in period of 1000 iteration, and all the 10000 sample points are used to estimate the densities. The Gibbs sampler seems to have reached convergence to its stationary distribution[5].

4.2 Empirical results

Our model without the autoregressive part seems to work rather well: a summary (means and percentiles) of the parameters' posteriors is shown in table 3, and the probabilities of the U.S. economy being in a contraction phase are plotted in figure 1.

Using the mean of the posterior distributions of the μ's as estimates for the growth rate, we obtain on a yearly base a contraction mean growth rate of -0.8% and an expansion mean growth rate of 5.1%.

In figure 1 the probabilities of the contraction state clearly discriminate the two phases of the economy, and the dating is similar to the one of the NBER (National Bureau of Economic Research).

[5] To save space we won't show graphs of the sample or diagnostics about convergence, but the sampled data are always available upon request.

	prior		posterior			
Parameter	mean	var.	mean	2.5%	50%	97.5%
μ_0	-0.50	1	-0.21	-0.62	-0.23	0.35
μ_1	1.50	1	1.47	1.14	1.46	1.81
β_1	0.00	5	1.53	0.64	1.53	2.45
β_2	0.00	5	-0.03	-0.32	-0.01	0.06
β_3	0.00	5	-2.75	-5.76	-2.62	-0.56
β_4	0.00	5	0.54	0.02	0.50	1.34
σ^2	-	∞	0.67	0.53	0.66	0.83

Table 3: Summary of the prior and posterior distributions of the parameters

Let's try now to answer the question that motivates the present paper: are the transition probabilities duration-dependent? The high concentration around zero of the posterior of the parameter β_2 seems to indicate that the probability of falling into a contraction is independent (or very weekly and vaguely dependent, if we also observe figure 3) of how long the economy has been in expansion. On the contrary the posterior of β_4 lays significantly away from zero, and figure 2 indicates a rather probable positive duration-dependence of the transition probability of moving into an expansion after a period of contraction.

5 Conclusion

In this paper it has been shown that it is possible to build a duration dependent Markov switching model remaining in the standard Markov switching framework. The application of such a model to the U.S. GDP data supports the hypothesis that there is significant duration dependence of the transition probability only when the economy is in a contraction state, but not vice versa. Our model expands the capabilities of Hamilton's Markov switching models and answers the criticism moved by Diebold et al. [7] to such models. Furthermore the Bayesian MCMC inference allows to simulate the joint distribution of the parameters, their marginal distributions and the distributions of possible transformations of them. The latent variable (state of the economy) is treated like a parameter and its posterior distribution is also simulated, as opposed to the ML framework, where only an inference conditional on the estimated parameters is possible.

Further work is needed to evaluate the forecasting performance of the model, and to verify whether the duration-dependance hypothesis is a feature common to the Business Cycle of other countries.

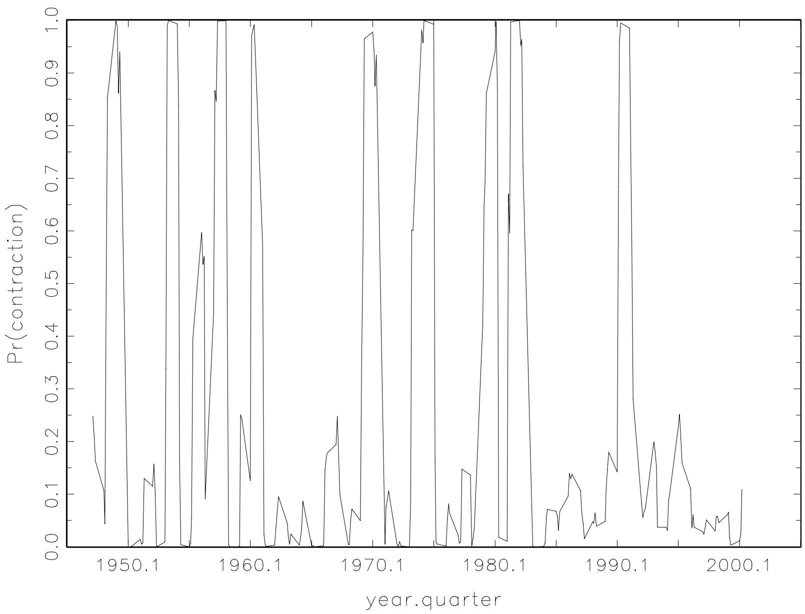

Figure 1: Probability of a contraction state for the U.S. economy

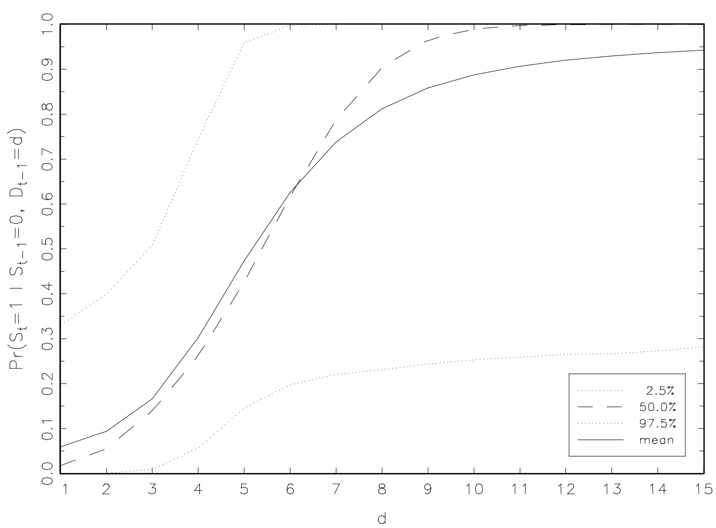

Figure 2: Probability of moving from a contraction to an expansion state after d quarters of contraction

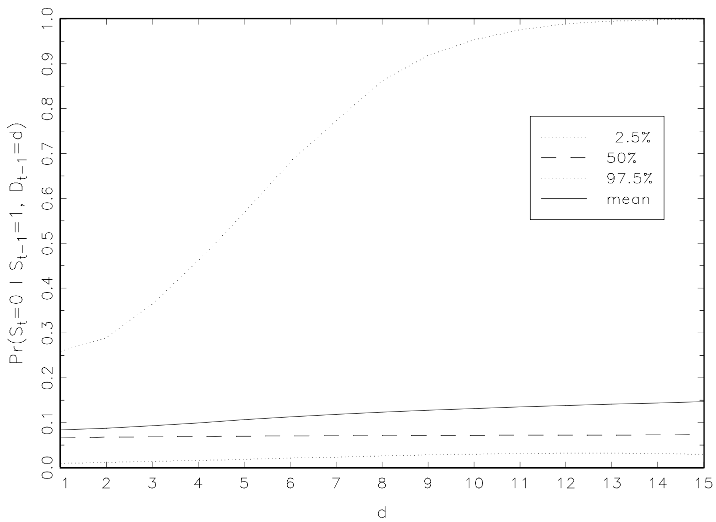

Figure 3: Probability of moving from an expansion state to a contraction state after d quarters of expansion

References

[1] J. H. Albert and S. Chib. Bayesian analysis of binary and polychotomous responce data. *Journal of the American Statistical Association*, 88:669–79, 1993.

[2] J. H. Albert and S. Chib. Bayesian inference via gibbs sampling of autoregressive time series subject to markov mean and variance shifts. *Journal of Business and Economic Statistics*, 1:1–15, 1993.

[3] B. P. Carlin, N. G. Polson, and D. S. Stoffer. A monte carlo approach to nonnormal and nonlinear state-space modeling. *Journal of the American Statistical Association*, 87:493–500, 1992.

[4] C. K. Carter and R. Kohn. On gibbs sampling for state space models. *Biometrika*, 81:541–53, 1994.

[5] C. K. Carter and R. Kohn. Markov chain monte carlo in conditionally gaussian state space models. *Biometrika*, 83:589–601, 1996.

[6] G. Casella and E. I. George. Explaining the gibbs sampler. *The American Statistician*, 46:167–173, 1992.

[7] F. Diebold, G. Rudebusch, and D. Sichel. Further evidence on business cycle duration dependence. In J. Stock and M. Watson, editors, Busi-

ness Cycles, Indicators and Forcasting. The University of Chicago Press, 1993.

[8] J. Durland and T. McCurdy. Duration–dependent transitions in a markov model of u.s. gnp growth. *Journal of Business and Economic Statistics*, 12:279–288, 1994.

[9] A. J. Filardo and S. F. Gordon. Business cycle durations. *Journal of Econometrics*, 85:99–123, 1998.

[10] A. E. Gelfand and A. F. M. Smith. Sampling–based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 89:398–409, 1990.

[11] J. D. Hamilton. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57:357–384, 1989.

[12] J. D. Hamilton. *Time Series Analysis*. Princeton University Press, 1994.

[13] L. Tierney. Markov chains for exploring posterior distributions. *The Annals of Statistics*, 22:1701–1762, 1994.

[14] J. Watson. Business cycle durations and postwar stabilization of the u.s. economy. *American Economic Review*, 84:24–46, 1994.

Appendix: Bayesian Gaussian Linear Models

In this section we just summarize few standard results for Bayesian linear models.

Let

$$\mathbf{y} \sim \mathcal{N}(\mathbf{X}\boldsymbol{\beta}, \sigma_y^2 \mathbf{I}_n) \tag{24}$$

be a linear model with, $\mathbf{y}$ a $n \times 1$ vector of observable quantities, $\mathbf{X}$ a $n \times k$ matrix of fixed regressors, $\boldsymbol{\beta}$ a $k \times 1$ vector of unknown parameters and σ_y^2 unknown variance of each element of $\mathbf{y}$.

If we model our prior believes about $\boldsymbol{\beta}$ and σ_y^2 in the well known (truncated) Normal-Gamma[6] form

$$\sigma_y^2 \sim \mathcal{IG}(n_0/2,\ n_0 v_0/2) \tag{25}$$

$$\boldsymbol{\beta}|\sigma_y^2 \sim \mathcal{N}(\mathbf{b}_0, \sigma_y^2 \mathbf{B}_0) I(\{\text{condition}\}) \tag{26}$$

[6] In formula $\mathcal{IG}$ denotes the inverse-gamma distribution

where $I(\{\text{condition}\})$ is an indicator variable that equals one only when the condition is true, and it is here used to denote truncation. After observing $\mathbf{y}$ we get the posterior distribution

$$\sigma_y^2 \sim \mathcal{IG}(n_1/2,\, n_1 v_1/2) \tag{27}$$

$$\boldsymbol{\beta}|\sigma_y^2 \sim \mathcal{N}(\mathbf{b}_1, \sigma_y^2 \mathbf{B}_1) I(\{\text{condition}\}) \tag{28}$$

with

$$\mathbf{B}_1 = (\mathbf{B}_0^{-1} + \mathbf{X}'\mathbf{X})^{-1} \tag{29}$$

$$\mathbf{b}_1 = \mathbf{B}_1(\mathbf{B}_0^{-1}\mathbf{b}_0 + \mathbf{X}'\mathbf{y}) \tag{30}$$

$$n_1 = n_0 + n \tag{31}$$

$$n_1 v_1 = n_0 v_0 + (\mathbf{y} - \mathbf{X}\mathbf{b}_1)'\mathbf{y} + (\mathbf{b}_0 - \mathbf{b}_1)'\mathbf{B}_0^{-1}\mathbf{b}_0 \tag{32}$$