Forecasting and Estimating Multiple Change-point Models

Gary Koop

University of Strathclyde

and

Simon Potter

Federal Reserve Bank of New York

1 Introduction

- Increasing recognition of importance of structural breaks (change-points) for empirical practice in macroeconomics and finance

- Structural breaks commonly cited as major source of forecasting failure (e.g. Clements and Hendry)

- Tests indicate structural breaks are widespread (e.g. Stock and Watson, 1996, Journal of Business and Economic Statistics)

- Huge literature on testing for structural breaks (e.g. Bai and Perron, 1998, Econometrica), less on developing models suitable for estimation and forecasting when structural breaks are present.

- Forecasting: how to model break process out-of-sample?

- How many breaks?

- Should we simply ignore pre-break data? Pastor and Stambaugh, 2001, Journal of Finance:

"In standard approaches to models that admit structural breaks, estimates of current parameters rely on data only since the most recent estimated break. ... That practice seems prudent, but it contends with the reality that shorter histories typically yield less precise estimates. Suppose.. a shift in the equity premium occured a month ago. ... Completely discarding the pre-break data is appropriate only when the premium might have shifted to such a degree that the pre-break data are no more useful .., than, say, pre-break rainfall data, but such a view almost surely ignores economics"

These considerations motivate our model which draws on our beliefs that desirable features for a change-point model are:

1. The parameters characterizing a new regime can potentially depend on the parameters of the old regime.

2. Durations of previous regimes can potentially provide some information about durations of future regimes.

3. The parameters describing the distribution of the parameters in each regime should, if possible, have conditionally conjugate prior distributions to minimize the computational complexity of change-point models.

4. The regime duration distribution should **not** be restricted to be constant or monotonically decreasing/increasing.

5. The number and maximum duration of regimes should **not** be restricted ex-ante.

6. Model should be able to nest small number of breaks up to T-1 breaks of a time varying parameter model.

2 Standard Approaches

Illustrate with regression model, but results general. All errors below N(0,1).

Model with Small Number of Breaks

$$
\begin{array}{l}
y_t = X_t\beta_1 + \sigma_1\varepsilon_t \text{ if } t \leq \tau_1 \\
y_t = X_t\beta_2 + \sigma_2\varepsilon_t \text{ if } \tau_1 < t \leq \tau_2 \\
y_t = X_t\beta_3 + \sigma_3\varepsilon_t \text{ if } t > \tau_2
\end{array}
$$

Terminology: here we have 3 regimes and, thus, two change-points τ_1 and τ_2

Equivalently can talk in terms of regime durations, $d_1 = \tau_1$, $d_2 = \tau_2 - \tau_1$ and $d_3 = T - \tau_2$

Equivalently can talk in terms of states: $s_t = 1$ if $t \leq \tau_1$, $s_t = 2$ if $\tau_1 < t \leq \tau_2$, etc.

In general have M regimes.

Issues:

- Pastor and Stambaugh issue. Few breaks, but completely unrestricted.

- How to forecast? Simply use coefficients in last regime. But what if more breaks occur?

- Computationally very challenging unless M is very small. E.g. Bai and Perron try every possible combination of breakpoints.

- M must be selected but bad consequences if you get it wrong.

Time-varying Parameter (TVP) Model

E.g. Cogley and Sargent (2001, 2003)

$$y_t = X_t \beta_t + \exp(\sigma_t/2)\varepsilon_t,$$

with

$$\begin{aligned} \beta_t &= \beta_{t-1} + \eta U_t, \\ \sigma_t &= \sigma_{t-1} + \omega u_t \end{aligned}$$

Issues:

- Have a break in every period.

- Coefficient changes across regimes constrained to be small.

- Note: can get around some problems by allowing η and ω to vary in some manner.

- This is a state space model and standard methods of estimation are available.

3 Bayesian Work: The Chib Model

Chib (1998), Journal of Econometrics popular in empirical work

Assumes s_t is Markovian. That is,

$$\Pr\left(s_t = j | s_{t-1} = i\right) = \begin{cases} p_i & \text{if } j = i \neq M \\ 1 - p_i & \text{if } j = i + 1 \\ 1 & \text{if } i = M \\ 0 & \text{otherwise} \end{cases}$$

Go from regime to regime. Once it has gone through the m^{th} regime, there is no returning to this regime. It goes through regimes sequentially, so it is not possible to skip from regime i to regime $i + 2$. Once it reaches the M^{th} regime it stays there.

Posterior computation can be done very efficiently drawing on MCMC algorithm of Chib (1996).

Let θ be model parameters, $S =$ vector of states. Works by sequentially drawing from:

$p\left(\theta|Data, S\right)$ and $p\left(S|Data, \theta\right)$

Issues:

- Fixed number of regimes.
- Pastor and Stambaugh issue. Few breaks, but completely unrestricted.
- How to forecast? Simply use coefficients in last regime. But what if more breaks occur?
- M must be selected but bad consequences if you get it wrong.

- Regime duration distribution is Geometric — decreasing probability.

- Pile-up of probability at end of sample.

Summary: Does not satisfy 5 out of 6 of our "desirable features of a change-point model"

4 A Poisson Hierarchical Model for Durations

We use a hierarchical prior for the regime durations which is a Poisson distribution.

$p\left(d_m|\lambda_m\right)$ is given by:

$$d_m - 1 = \tau_m - (\tau_{m-1} + 1) \sim Po(\lambda_m)$$

Issues

- Does not impose a fixed number of regimes (some can occur out-of-sample)

- Poisson is commonly-used flexible distribution. Computation (relatively) straightforward.

- Unlike other approaches, the regime duration distribution is **not** be restricted to be constant or monotonically decreasing/increasing.

Remember, we also wanted:

"Durations of previous regimes can potentially provide some information about durations of future regimes."

We do this through another hierarchical prior of the form:

$$\lambda_m \,|\beta_\lambda \sim G\left(\underline{\alpha}_\lambda, \beta_\lambda\right),$$

where β_λ is an unknown parameter.

λ_m controls the duration of the m^{th} regime.

We are saying this is drawn from some common distribution estimated from the data. Information from all regimes is used to estimate this distribution (i.e. estimate β_λ).

Duration of out-of-sample regimes depends on data (i.e. data used to estimate β_λ). Key for forecasting.

5 Development of the Prior for the Parameters in Each Regime

We assume, for $m = 1, .., M$ regimes

$$y_t = X_t\beta_m + \exp(\sigma_m/2)\varepsilon_t,$$

$$\begin{aligned}\beta_m &= \beta_{m-1} + U_m, \\ \sigma_m &= \sigma_{m-1} + u_m,\end{aligned}$$

Like the TVP model this is a state space model, but non-standard. Note the subscripts.

This satisfies:

"Model should be able to nest small number of breaks up to T-1 breaks of a time varying parameter model"

We assume $U_m \sim N(0, V)$, $u_m \sim N(0, \eta)$.

V and η controls size of coefficient change across regimes.

Note: Simple extension: all V and η to vary across regimes.

Aside: for V_m we can use yet another hierarchical prior:

$$V_m^{-1} \sim W\left(\frac{\left[\lambda_{m-1}\mathcal{V}_V\right]^{-1}}{\underline{\nu}_V - K - 1}, \underline{\nu}_V\right)$$

where $W(A, a)$ is the Wishart distribution. $\mathcal{V}_V$ and $\mathcal{V}_\eta$ are unknown parameters estimated from the data.

Note: V_m drawn from some common distribution. These depend on λ_{m-1} – so size of coefficient change across regimes can depend on duration of previous regime.

With all these, we now have a model satisfying:

"The parameters characterizing a new regime can potentially depend on the parameters and duration of the old regime." May be useful for forecasting.

5.1 Estimation, Testing and Prediction in Poisson Hierarchical Model

Markov Chain Monte Carlo Algorithm that modifies well established algorithms:

- States, s_t, drawn using a modified version of Chib (1996)

- Regression coefficients drawn using (modified) algorithms for state space models.

- Error variances drawn using algorithm for stochastic volatility model.

- Poisson intensities (λ_m) standard results for Poisson likelihoods, except for incomplete regime (see paper for details).

- Other parameters – see paper (simple forms).

- Giordani and Kohn (2006) algorithm is more efficient (and could be adapted to our model).

- Paper contains a real time forecasting exercise and shows how sequential importance sampling (particle filtering) methods can be helpful for this (i.e. MCMC algorithm run once, rather than run for each real time forecasting point).

6 Application to US Inflation and Output

US data from 1947Q1 through 2004Q2

Issue 1: Declining volatility of GDP growth usually dated to mid 1980s.

Issue 2: Changing persistent and volatility of inflation.

Compare our model to TVP model and a one-break model.

In general, our model yields results between these two, but closer to TVP.

See figures.

Table 1: Joint Predictive Likelihoods for 2004/2005

	Our Model	TVP	One Break
GDP Growth	9.77×10^{-7}	6.06×10^{-7}	4.93×10^{-7}
Inflation	3.69×10^{-6}	3.31×10^{-6}	1.75×10^{-6}

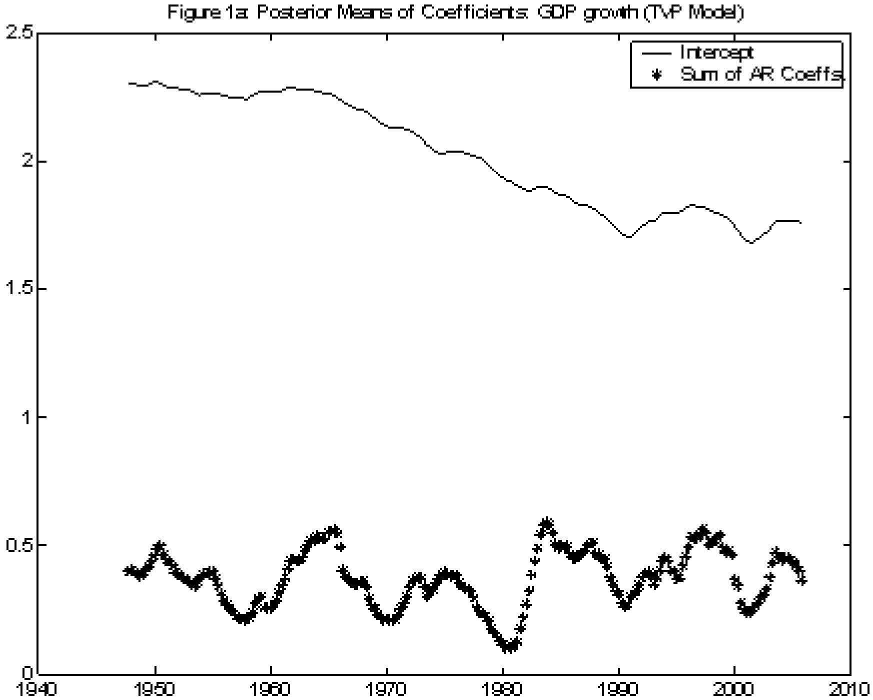


Figure 1a: Posterior Means of Coefficients: GDP growth (TVP Model)
Intercept
Sum of AR Coeffs.
2.5
2
1.5
1
0.5
0
1940
1950
1960
1970
1980
1990
2000
2010

Figure 1b: Posterior Mean of Volatility: GDP growth (TvP Model)

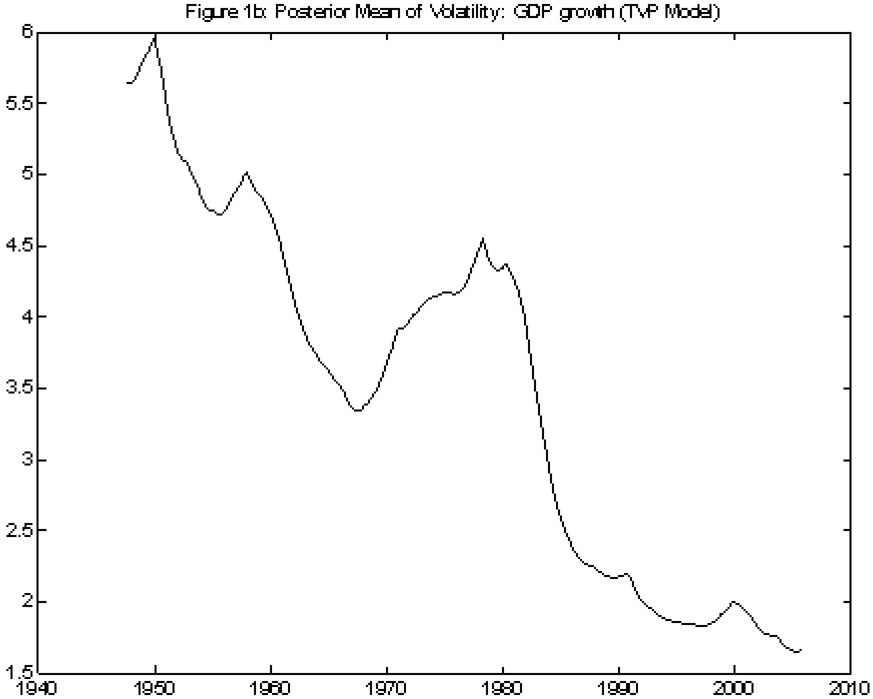


Figure 2a: Posterior Means of Coefficients: GDP Growth (One Break Model)

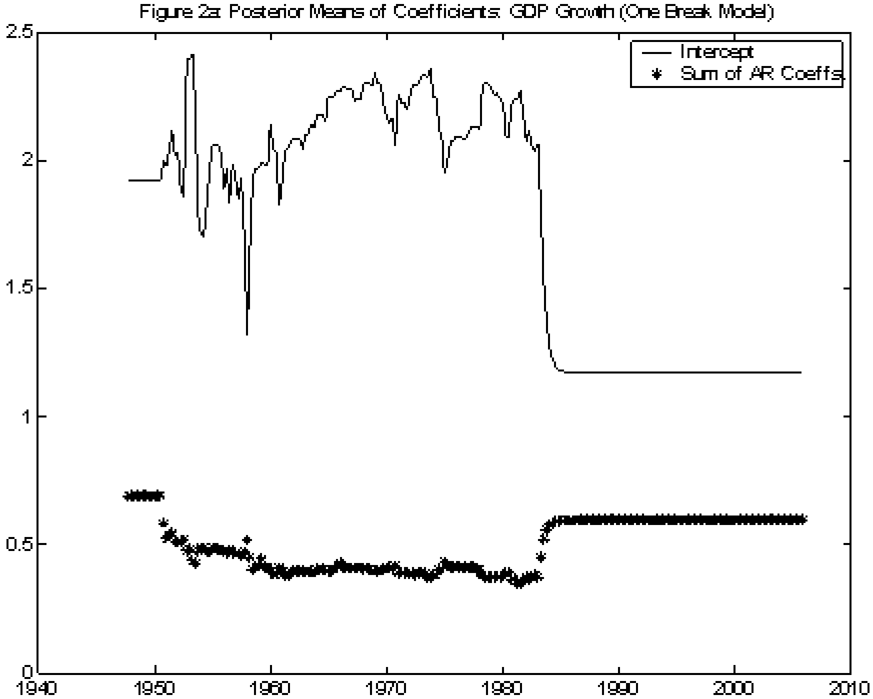


Figure 2b: Posterior Mean of Volatility: GDP Growth (One Break Model)

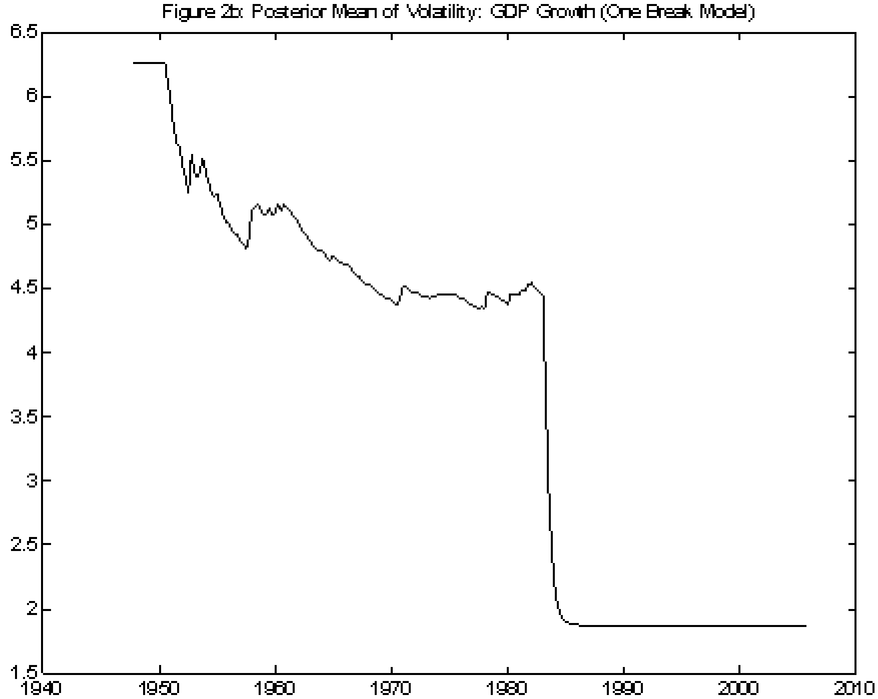
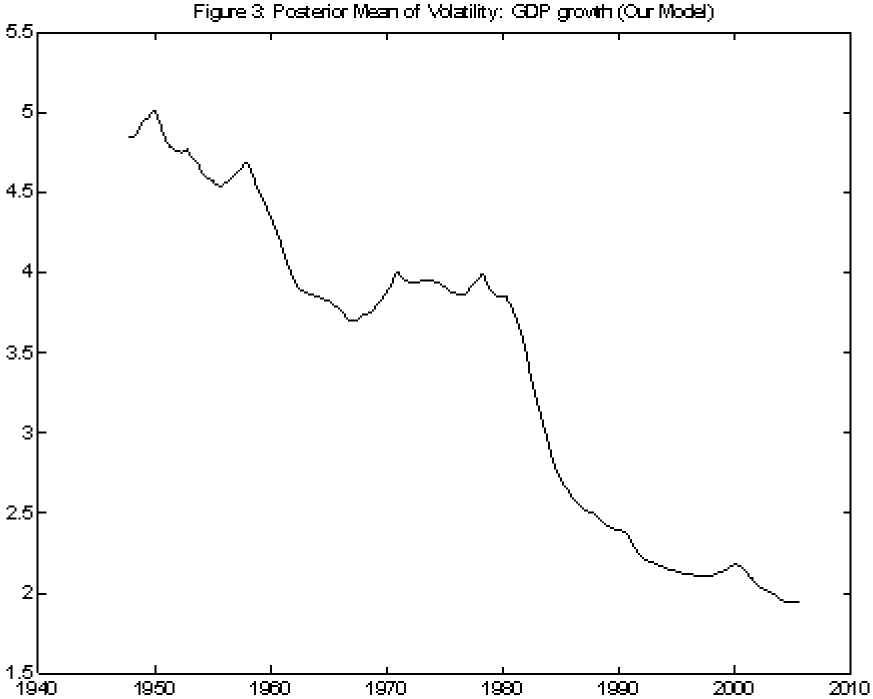


Figure 3: Posterior Mean of Volatility: GDP growth (Our Model)



7 Conclusions

We have derived a change-point model (and Bayesian methods for estimation, testing and prediction) which satisfies our desirable criteria:

1. The parameters characterizing a new regime can potentially depend on the parameters of the old regime.

2. Durations of previous regimes can potentially provide some information about durations of future regimes.

3. The parameters describing the distribution of the parameters in each regime should, if possible, have conditionally conjugate prior distributions to minimize the computational complexity of change-point models.

4. The regime duration distribution should **not** be restricted to be constant or monotonically decreasing/increasing.

5. The number and maximum duration of regimes should **not** be restricted ex-ante.

6. Model should be able to nest small number of breaks up to T-1 breaks of a time varying parameter model.

Such a model is motivated by desire for precise estimation, forecasting and efficient computation.

Figure 4a: Posterior Means of Coefficients: Inflation (TVP Model)

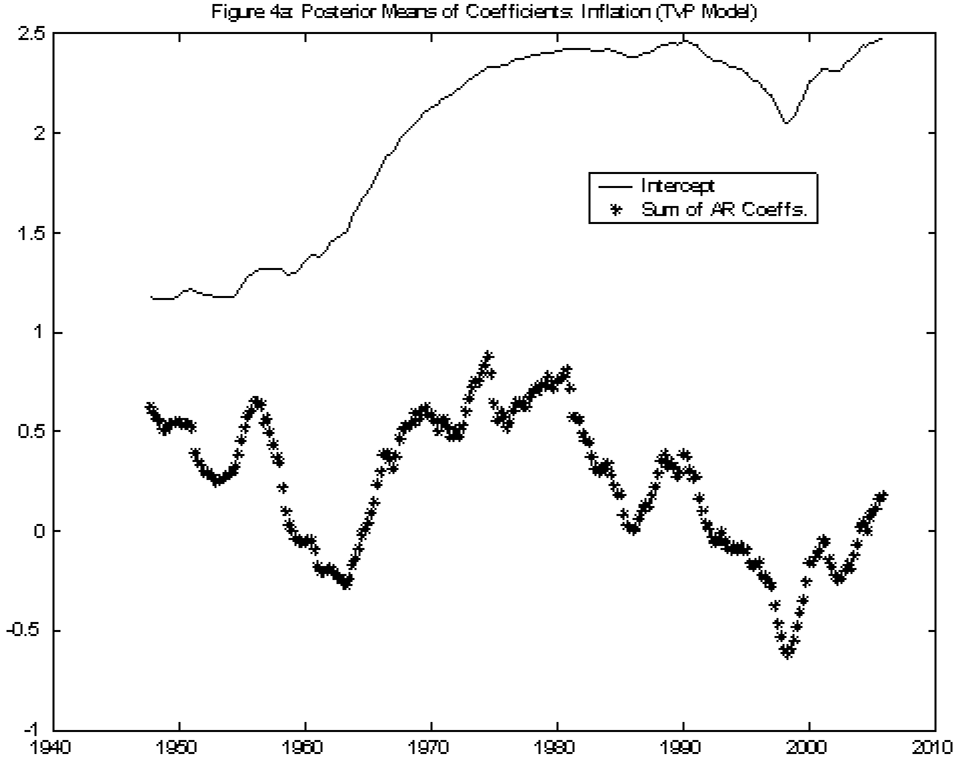


Figure 4b: Posterior Mean of Volatility: Inflation (TVP Model)

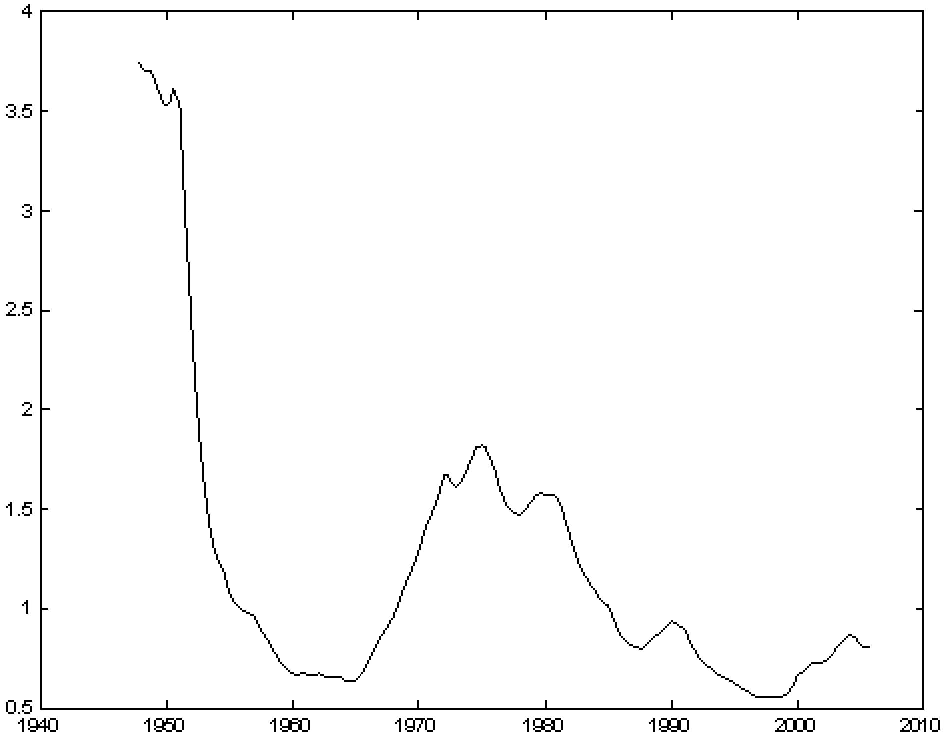


Figure 5a: Posterior Means of Coefficients: Inflation (One Break Model)

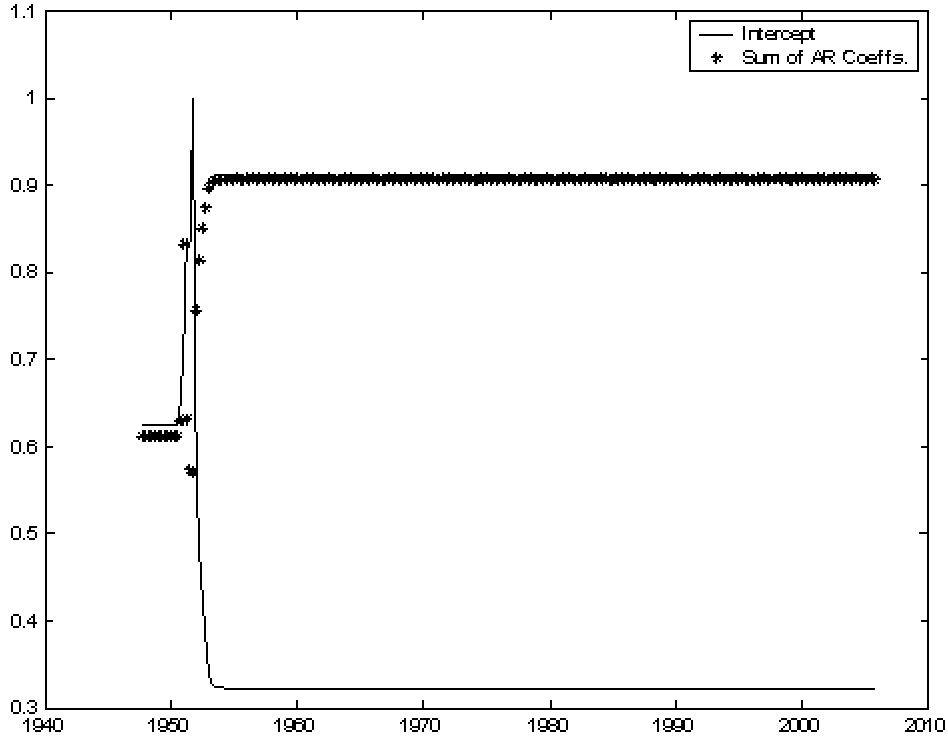


Figure 5b: Posterior Mean of Volatility: Inflation (One Break Model)

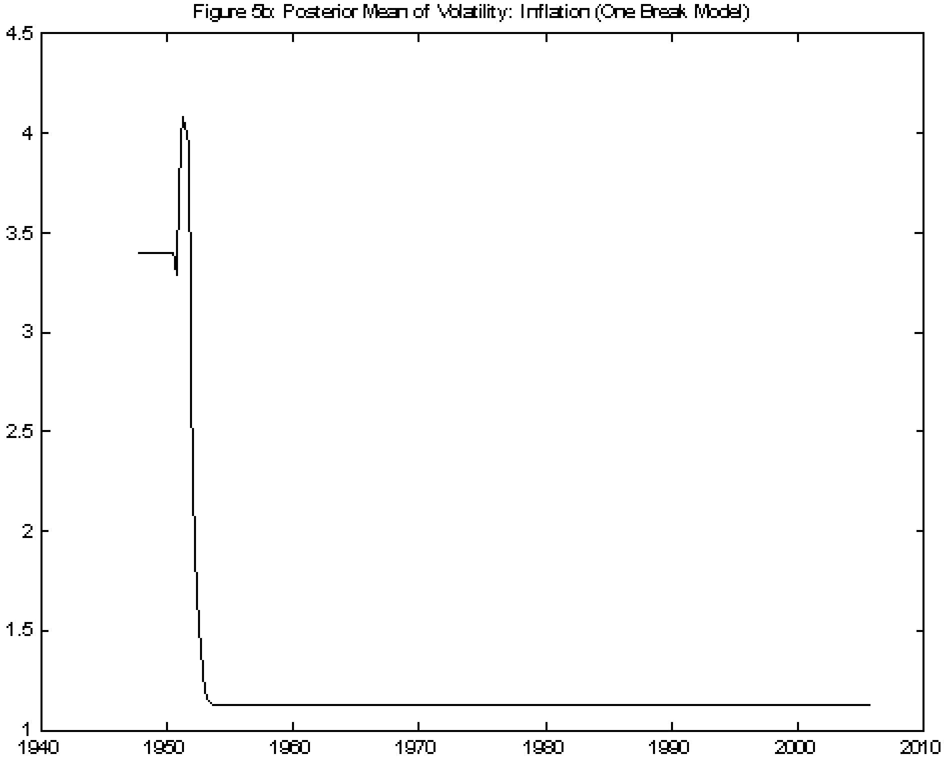


Figure 6a: Posterior Means of Coefficients: Inflation (Our Model)

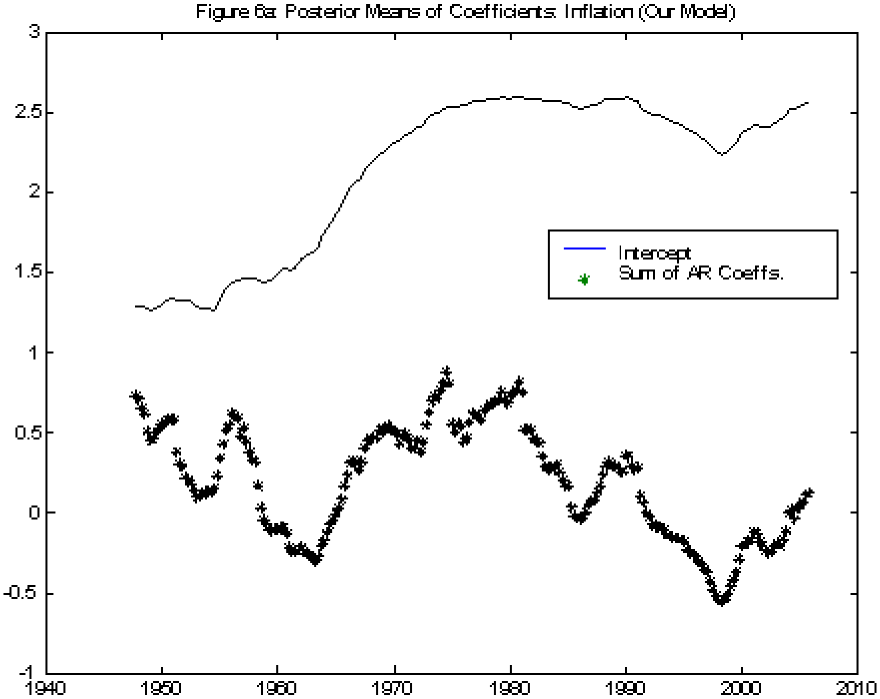


Figure 6b: Posterior Mean of Volatility: Inflation (Our Model)