Estimation and Forecasting in Models with Multiple Breaks

Gary Koop and Simon Potter

University of Strathcylde and New York Fed

November 2006

The views expressed are not necessarily those of FRBNY or FRS

Examples of structural change/parameter variation

- GDP and nearly every other macroeconomic series are less volatile 1984 to present than before
- Well-known (reduced form) relationships used by central banks vary over time
  1. Okun's Law
  2. Phillips Curve
- Pervasiveness of unit roots, near unit roots, fractional integration in economic time series
- Mix of production/employment/consumption changed greatly over the last 100 years
- Individuals live longer, better educated, fewer children

Motivation

Why do we care?
For purposes here take a narrow technical definition:
Observed time series do not have one of the following

- linear form (ie IID errors)
- Fixed parameters
- Initial conditions drawn from its stationary distribution.

Forecasting

If constant parameter linear iid true

- More accurate model choice and parameter estimates (97% of econometrics) $\longrightarrow$ Better Forecasts
- In practice no evidence of better forecasting by economists:
  1. Despite massive improvements in econometric techniques under constant parameter linear iid restriction .
  2. Various papers on forecasting by Stock and Watson conclude parameter variation of subtle type is present
- Forecasting Models need to take account of structural change/parameter instability
  - In particular they need to be able to “deal” with change in the recent past and the certainty of more change in the future

Not just Lucas Critique

- A Major (but not only) source of parameter instability is failure to analyze "structural relationships/parameters"
- Usual (heroic) assumption is that "structural parameters" are constant, like the Rocky Mountains.
- Changes in life expectancy, education, knowledge and fertility happen at a much higher frequency than changes in the Rock Mountains
- Empirically modeling these changes as part of the transition to the long run stochastic steady state virtually impossible
- Lots of recent work on policymaking under uncertainty
- **Assertion**: policymakers do better if forecasting models constructed to allow for parameter instability
  - At least gives some method of analyzing whether they are moving things in the desired direction

Literature Review

- 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? see Pesaran et al ReStud forthcoming
- Should we simply ignore pre-break data? Pastor and Stambaugh, 2001, Journal of Finance

Desirable Properties for multiple break model

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.

Model with Small Number of Breaks

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

Terminology: here we have 3 regimes and, thus, two change-points τ_1 and τ_2 or in terms of regime durations, $d_1 = \tau_1$, $d_2 = \tau_2 - \tau_1$ and $d_3 = T - \tau_2$ or 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.

- Few breaks, but completely unrestricted
- How to forecast? Simply use coefficients in last regime. But what if more breaks occur out of sample?
- Computationally very challenging unless M is very small
- M must be selected but bad consequences if you get it wrong

Time-varying Parameter (TVP) Model

E.g. Cogley and Sargent (2001, 2005,2006)

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

with

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

- Break in every period
- Coefficient changes across regimes constrained to be small
- 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

Chib Model

Chib (1998), Journal of Econometrics popular in empirical work
Assumes s_t is Markovian. That is,

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

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

Issues

- Regime duration distribution is Geometric — decreasing probability.
- Pile-up of probability at end of sample.

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

Uniform priors on duration

Chib's Markov switching formulation is interpreted by Bayesians as a "hierarchical prior". An alternative, in the spirit of the non-Bayesian literature:

Restricted Uniform Prior

(e.g. two change-point case)

$$p(\tau_1, \tau_2) = p(\tau_1)\, p(\tau_2 | \tau_1)$$

$$p(\tau_1) = \frac{1}{T-2} \text{ for } \tau_1 = 1, .., T-2$$

$$p(\tau_2 | \tau_1) = \frac{1}{T - \tau_1 - 1} \text{ for } \tau_2 = \tau_1 + 1, .., T-1.$$

Has similar problems to Chib's prior.

Unrestricted Uniform Prior

$$p(\tau_1) = \frac{1}{T-2} \text{ for } \tau_1 = 1, .., T-2.$$

$$p(\tau_2|\tau_1) = \frac{1}{T-2} \text{ for } \tau_2 = \tau_1 + 1, .., T + \tau_1 - 2.$$

Solves undesirable properties of Chib's and Restricted Uniform prior, but allocates probability to breaks occurring out-of-sample! This is actually a desirable property since it implicitly allows for unknown number of change-points. E.g. if τ_2 occurs out-of-sample we have a one change-point model.
Note this also allows us to forecast out-of-sample breaks.

Enriched Unrestricted Uniform Prior

For the case where there is a maximum of $M-1$ breaks in-sample, we write our prior as $p\left(\tau_1,\tau_2,..,\tau_{M-1}\right)=p\left(\tau_1\right)\sum_{j=2}^{M-1}p\left(\tau_j|\tau_{j-1}\right)$ and assume

$$p\left(\tau_1\right)=\frac{1}{[cT]}\text{ for }\tau_1=1,..,[cT].$$

and

$$p\left(\tau_j|\tau_{j-1}\right)=\frac{1}{[cT]}\text{ for }\tau_j=\tau_{j-1}+1,..,\tau_{j-1}+[cT]$$

The notation $[cT]$ indicates the smallest integer such that $cT\leq[cT]$.
If $c=\frac{1}{T}$ we obtain the TVP model.
Previous example set $M=2$ and set $c=\frac{T-2}{T}$.

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

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

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

Issues

- Does not impose a fixed number of regimes (some can occur out-of-sample)
- Poisson is commonly-used flexible distribution. Computation straightforward
- Unlike other approaches, the regime duration distribution is **not** be restricted to be constant or monotonically decreasing/increasing

Heirarchy

"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.

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

Estimation 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 earlier version of paper on Gary's website (simple forms).

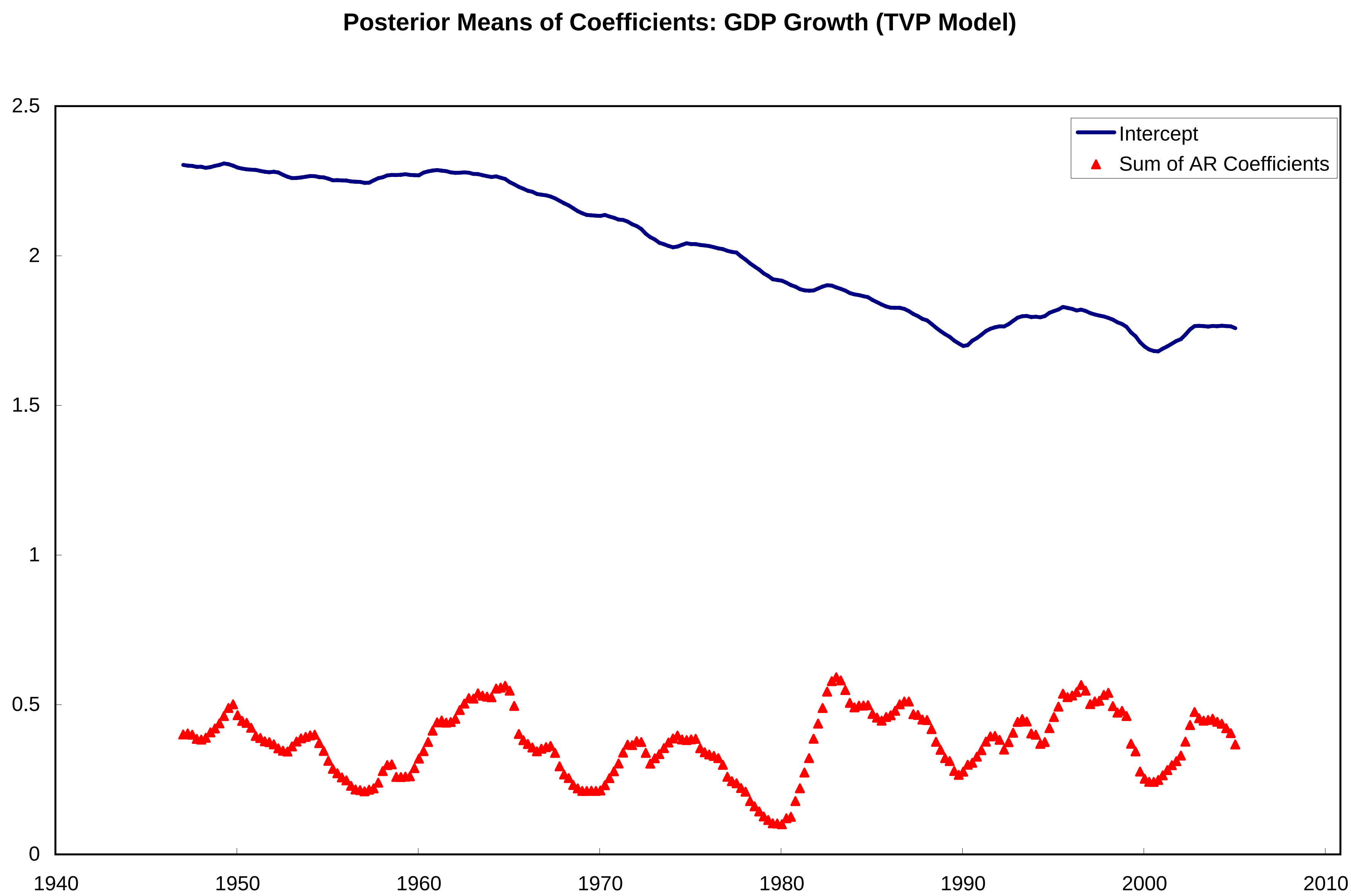
Application to US Inflation and Output

US data from 1947Q1 through 2005Q4

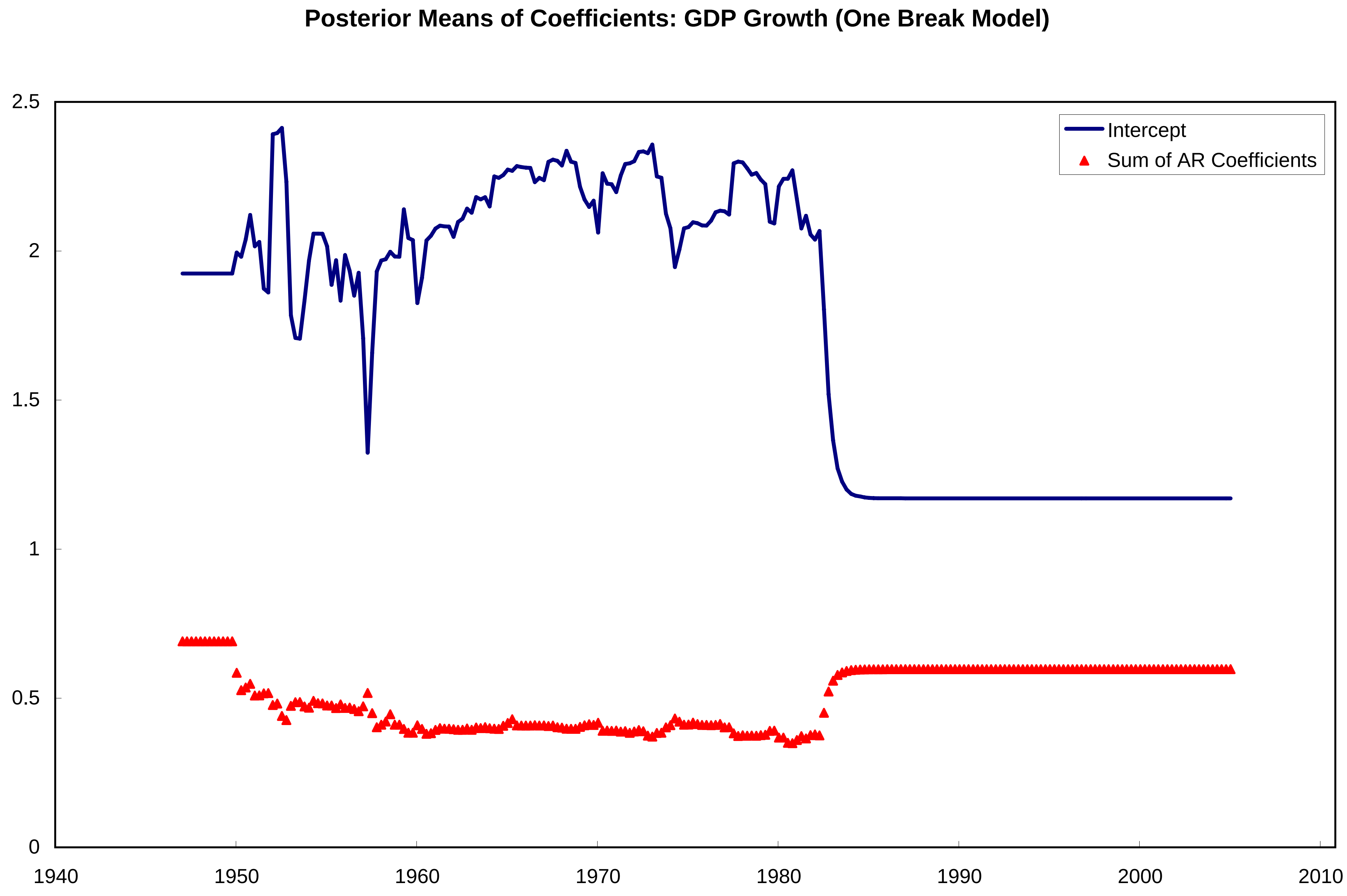
- Real GDP growth
- Inflarion: PCE deflator

1. Declining volatility of GDP growth usually dated to mid 1980s.
2. Changing persistent and volatility of inflation.

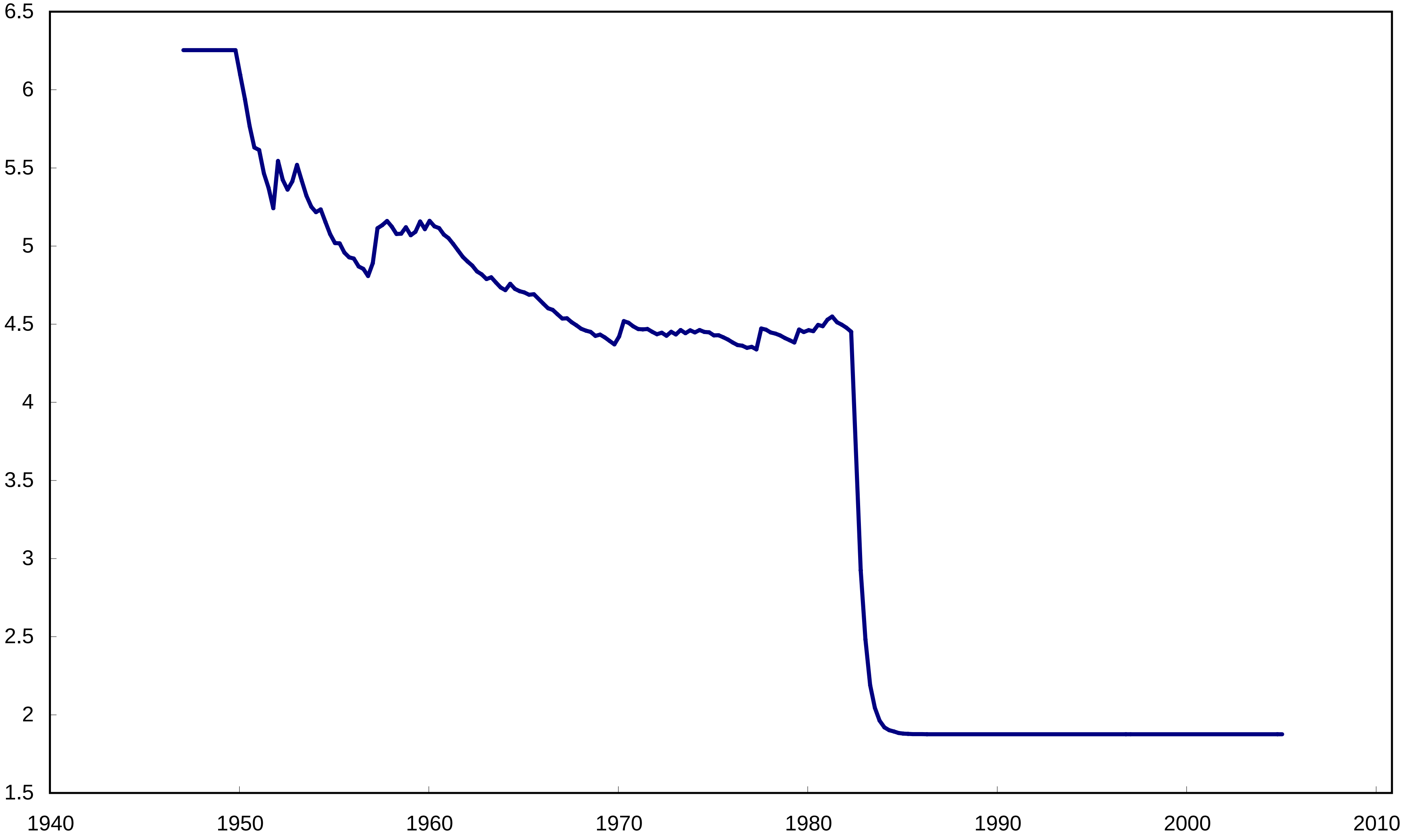
Compare our model to TVP model and a one-break model.
In general, our model yields results between these two, but closer to TVP.

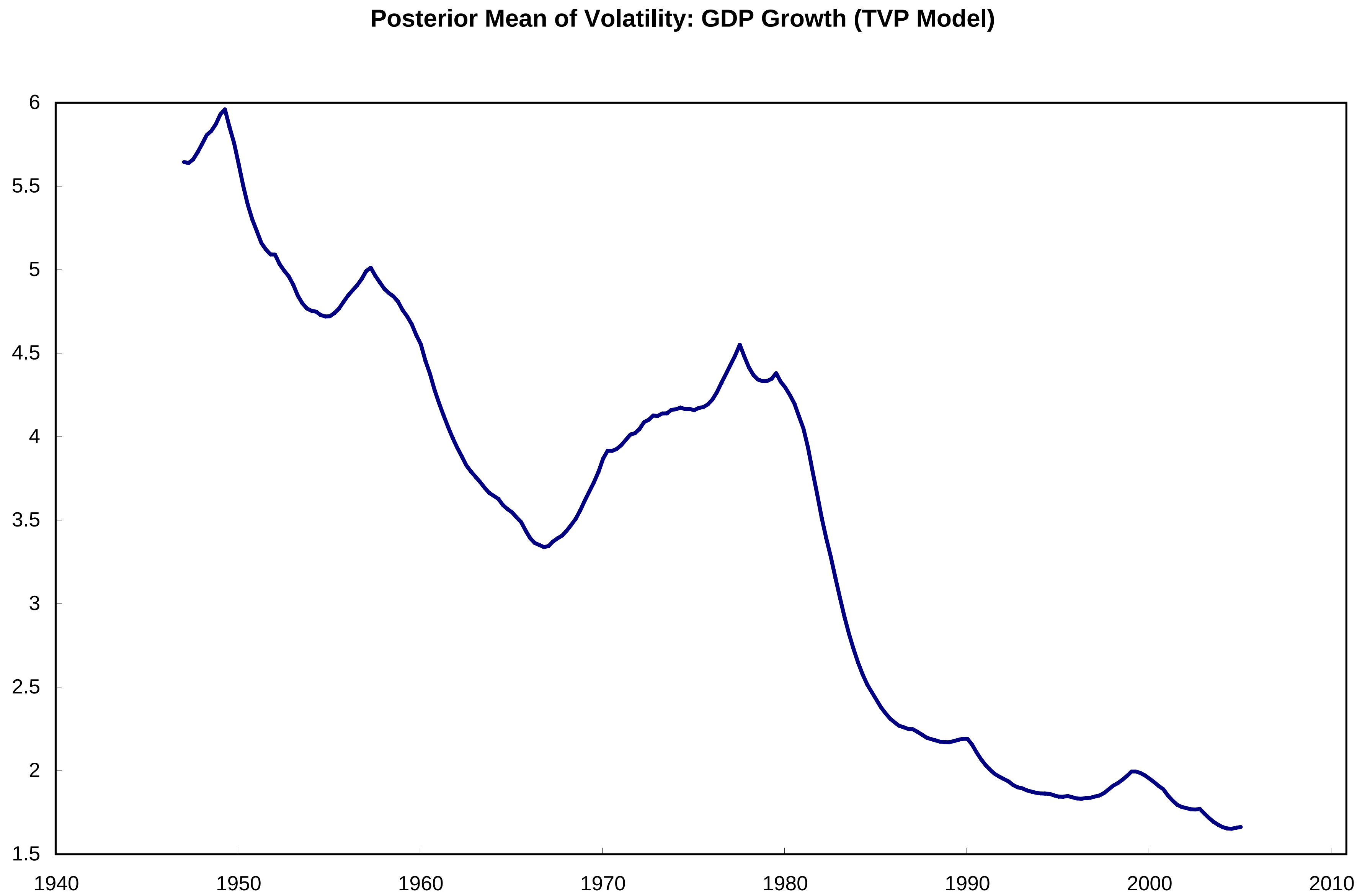


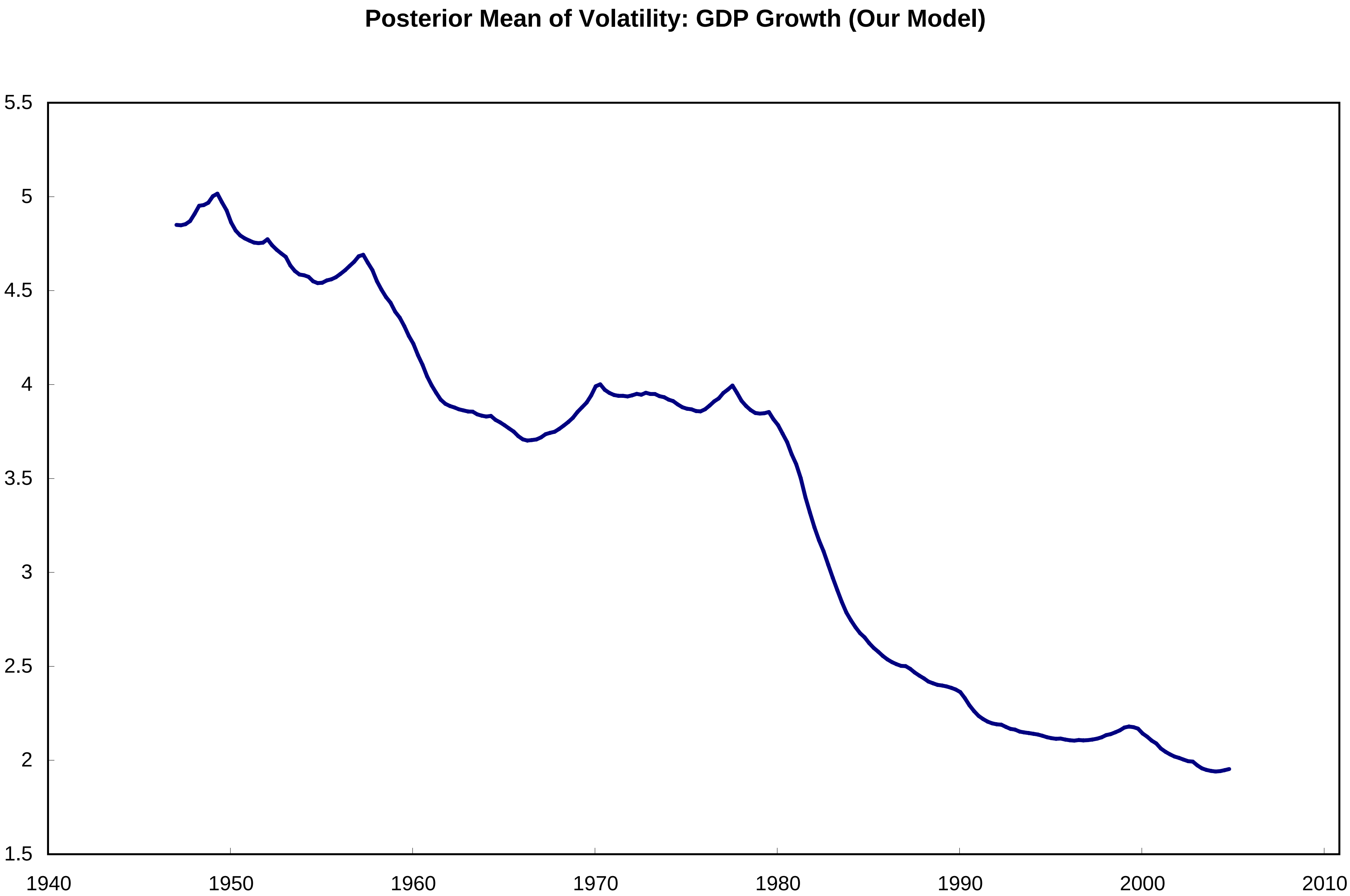
Posterior Means of Coefficients: GDP Growth (TVP Model)
2.5
2
1.5
1
0.5
0
1940
1950
1960
1970
1980
1990
2000
2010
Intercept
Sum of AR Coefficients


Posterior Means of Coefficients: GDP Growth (One Break Model)
2.5
2
1.5
1
0.5
0
1940
1950
1960
1970
1980
1990
2000
2010
Intercept
Sum of AR Coefficients

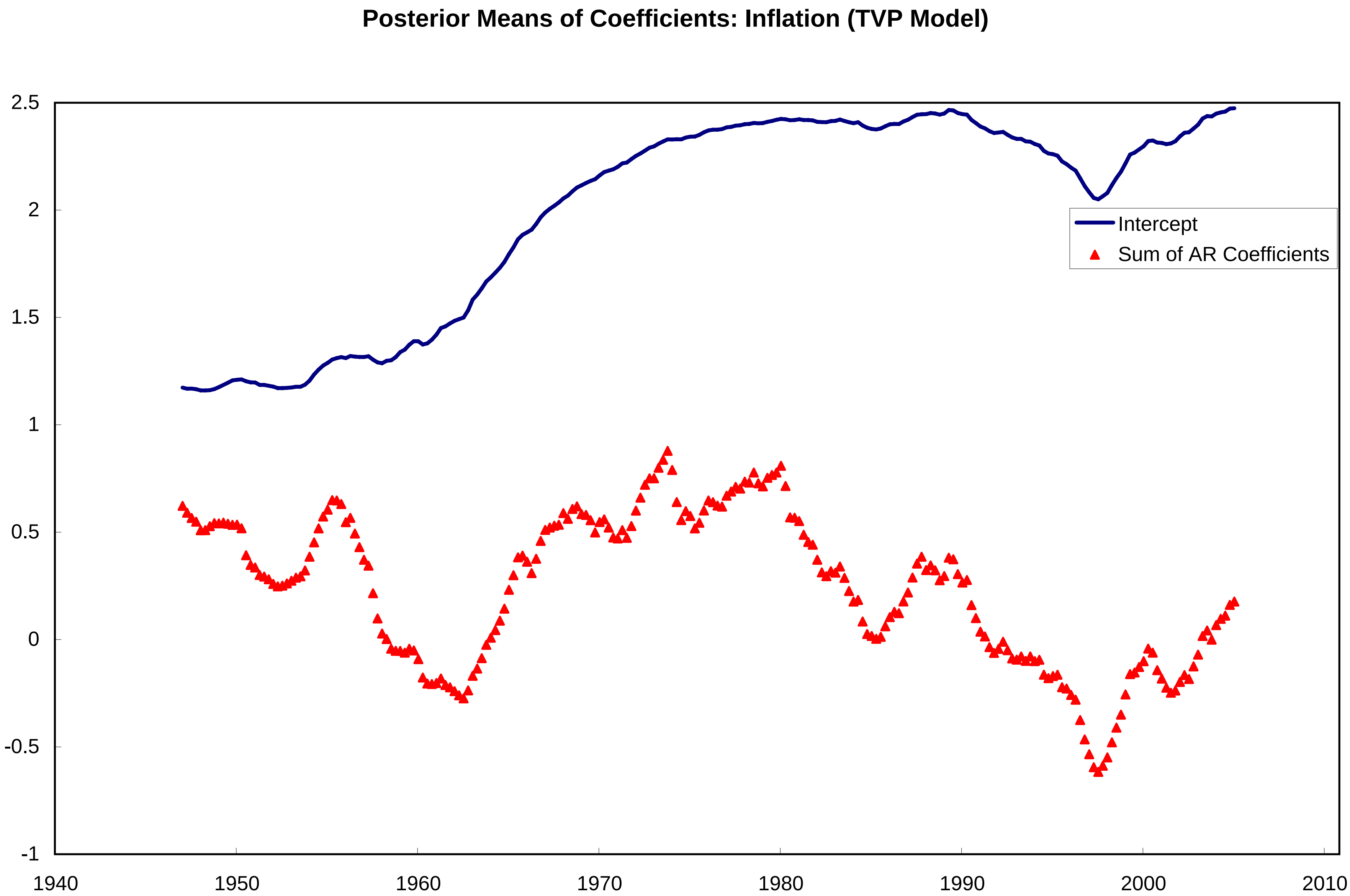
Posterior Mean of Volatility: GDP Growth (One Break Model)

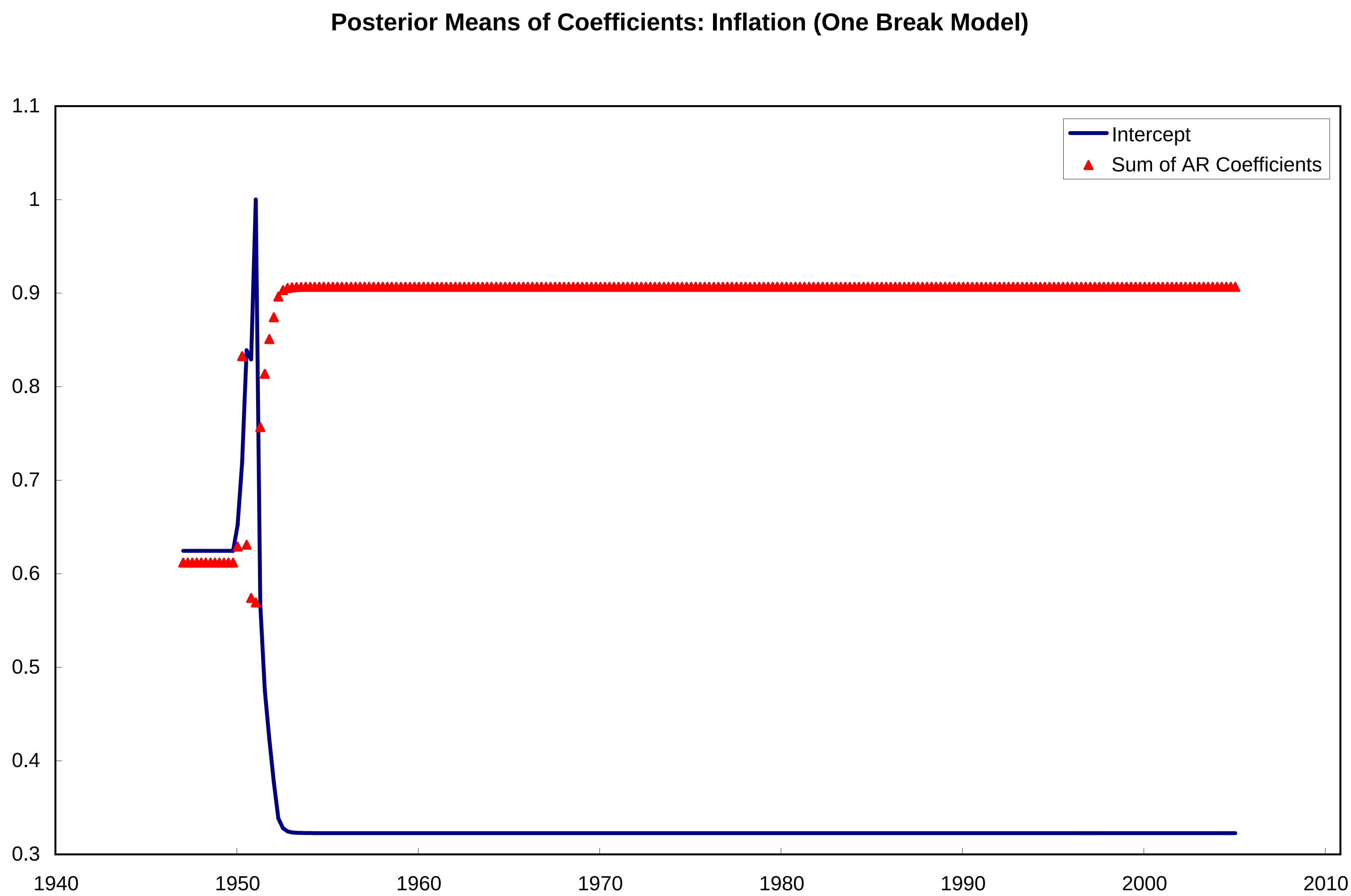


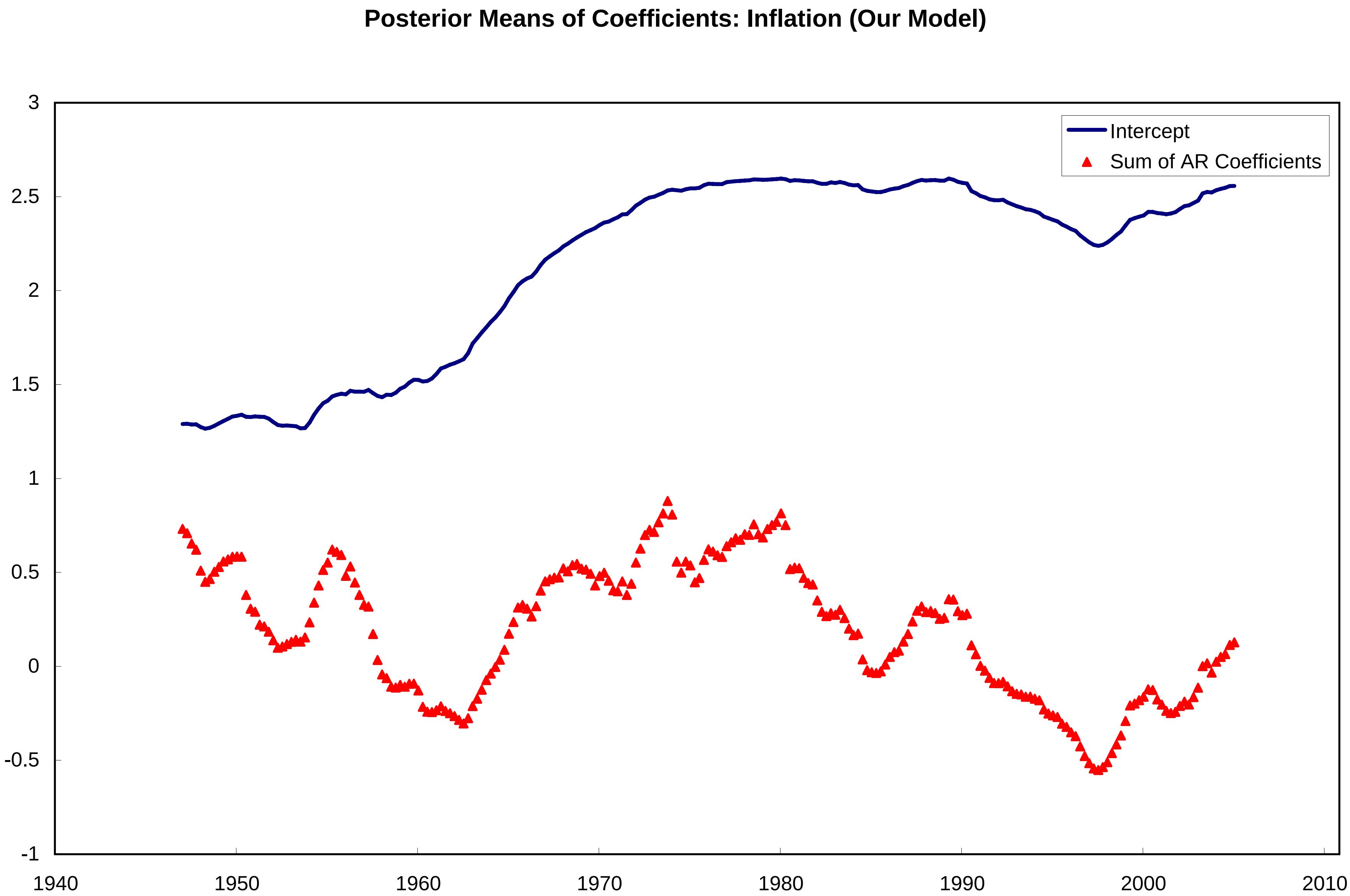

Posterior Mean of Volatility: GDP Growth (TVP Model)
6
5.5
5
4.5
4
3.5
3
2.5
2
1.5
1940
1950
1960
1970
1980
1990
2000
2010

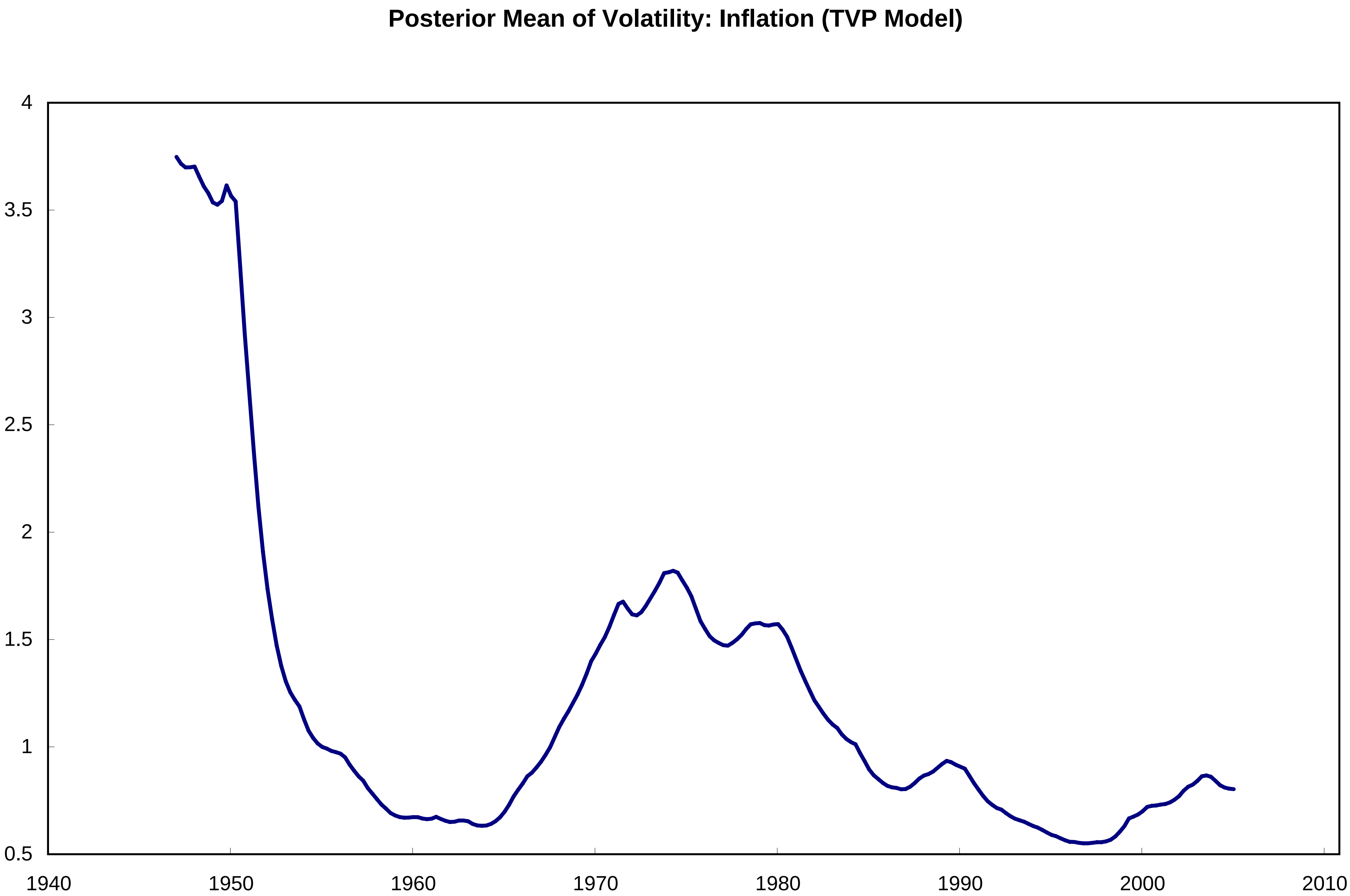


Posterior Mean of Volatility: GDP Growth (Our Model)
5.5
5
4.5
4
3.5
3
2.5
2
1.5
1940
1950
1960
1970
1980
1990
2000
2010

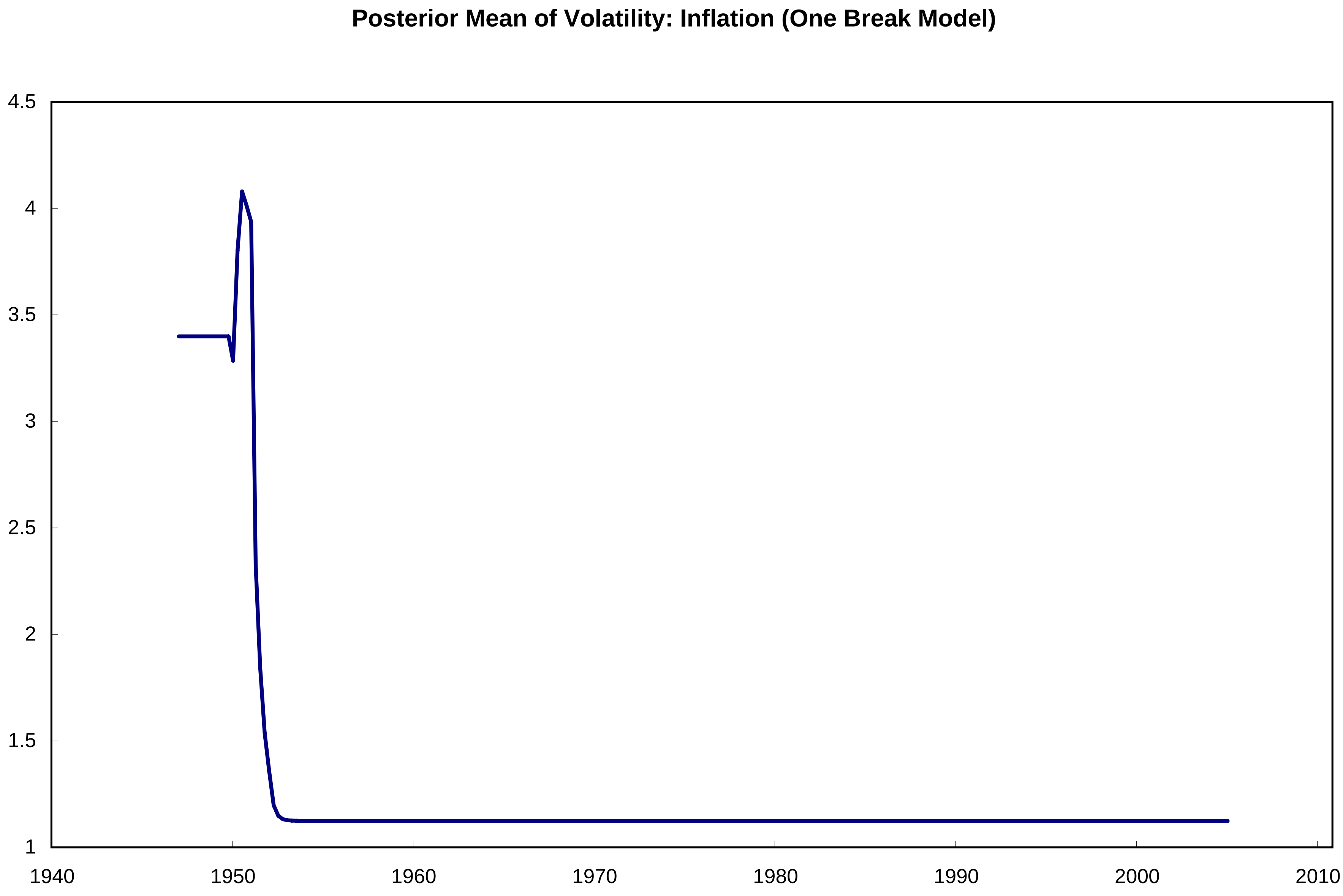

Posterior Means of Coefficients: Inflation (TVP Model)
2.5
2
1.5
1
0.5
0
-0.5
-1
1940
1950
1960
1970
1980
1990
2000
2010
Intercept
Sum of AR Coefficients


Posterior Means of Coefficients: Inflation (One Break Model)
Intercept
Sum of AR Coefficients
1.1
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
1940
1950
1960
1970
1980
1990
2000
2010


Posterior Means of Coefficients: Inflation (Our Model)
Intercept
Sum of AR Coefficients
3
2.5
2
1.5
1
0.5
0
-0.5
-1
1940
1950
1960
1970
1980
1990
2000
2010



Posterior Mean of Volatility: Inflation (TVP Model)
4
3.5
3
2.5
2
1.5
1
0.5
1940
1950
1960
1970
1980
1990
2000
2010


Posterior Mean of Volatility: Inflation (One Break Model)
4.5
4
3.5
3
2.5
2
1.5
1
1940
1950
1960
1970
1980
1990
2000
2010

Posterior Mean of Volatility: Inflation (Our Model)

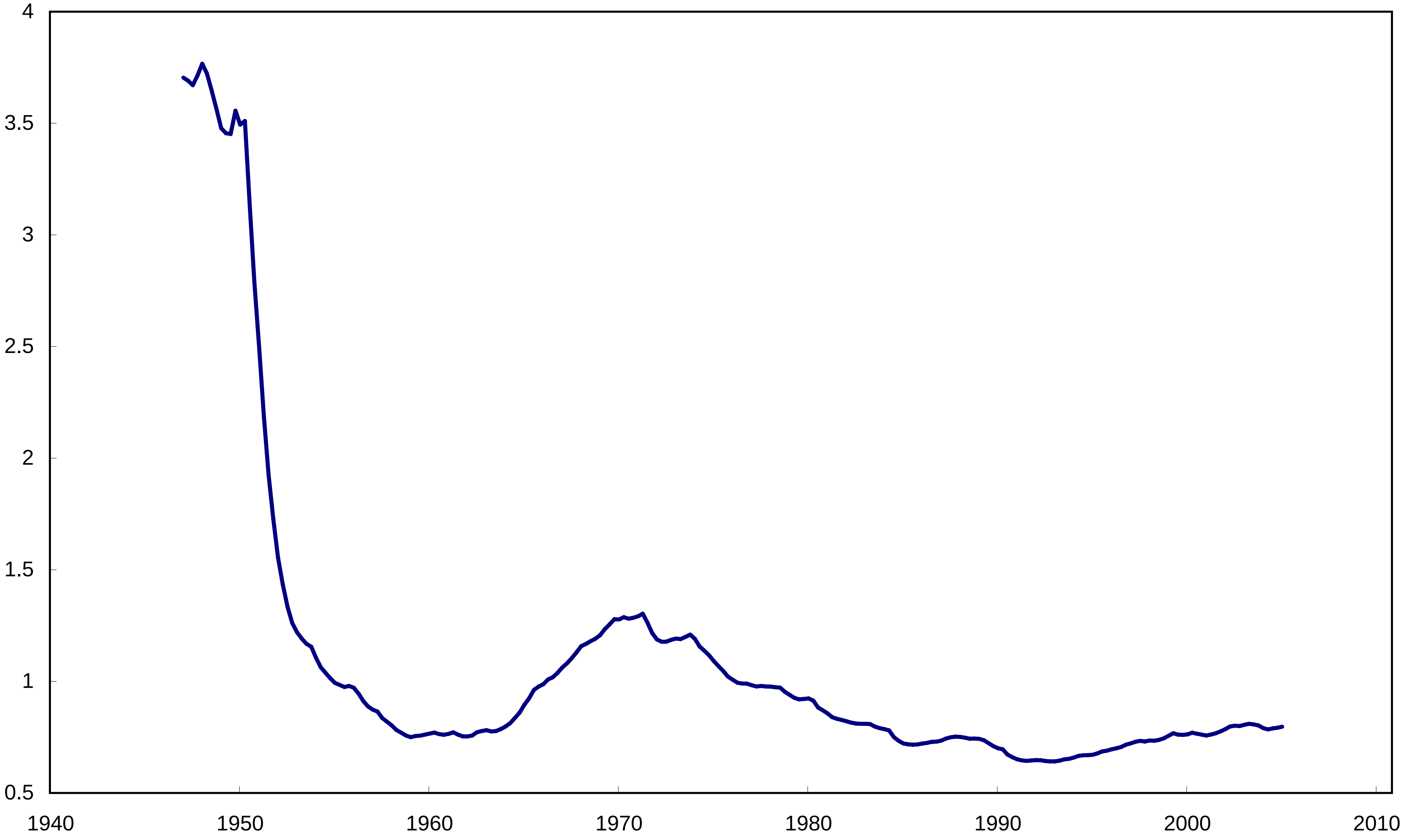


Figure 7a: Predictive for 2006Q1 (GDP Growth)

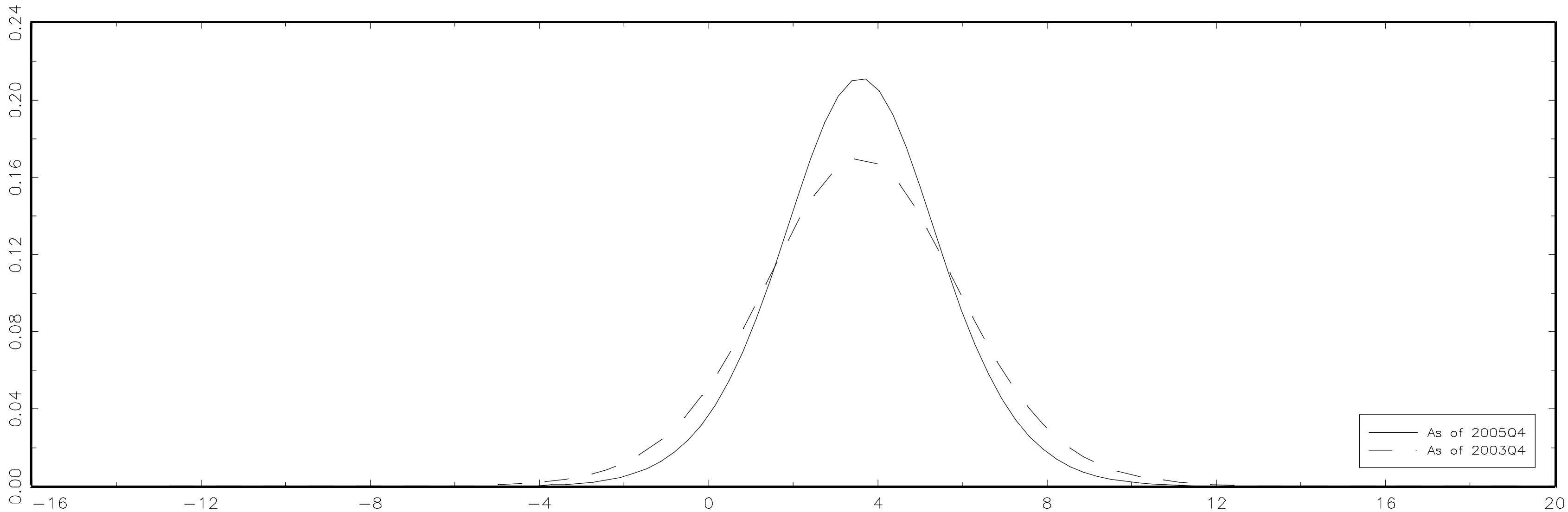
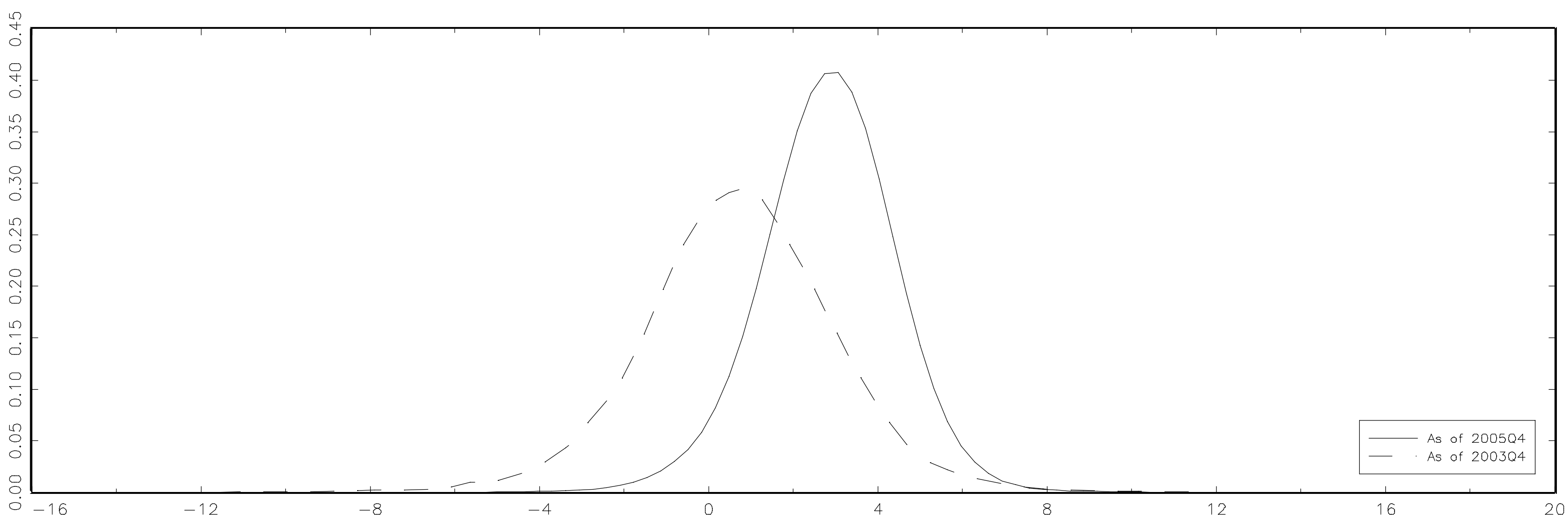


Figure 7b: Predictive for 2006Q1 (Inflation)



Number of "Regimes" for GDP and Prior Sensitivity

Prior Sensitivity Analysis For GDP Growth			
Posterior Mean of # Regimes			
	$\underline{\xi}_2 = 1$	$\underline{\xi}_2 = 12$	$\underline{\xi}_2 = 100$
$\underline{\xi}_1 = 1$	73.32	12.48	12.15
$\underline{\xi}_1 = 12$	74.16	45.35	12.17
$\underline{\xi}_1 = 100$	78.29	76.23	17.38

Number of "Regimes" for Inflation and Prior Sensitivity

Prior Sensitivity Analysis For Inflation			
Posterior Mean of # Regimes			
	$\underline{\xi}_2 = 1$	$\underline{\xi}_2 = 12$	$\underline{\xi}_2 = 100$
$\underline{\xi}_1 = 1$	189.25	90.32	86.96
$\underline{\xi}_1 = 12$	189.80	124.00	87.96
$\underline{\xi}_1 = 100$	190.62	186.85	88.92

Model contains 6 desirable features

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 multiple 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.

Further Research

- Improved Algorithms
  1. Paper uses a "particle filter" to ease real time updating
  2. Adaptive methods Metropolis Hastings: potential big improvements (Giordani and Kohn)
- Model variations
  1. Full version of model (regime parameters depend on poisson intensity) for financial time series
  2. New paper with Koop, averages across nonlinear and break models
  3. Bound the variation in parameters a la Cogley and Sargent
     - Interesting when applied to natural rate models/unobserved component models