Bayesian Econometrics of Learning Models (lecture notes for the Dynare learning workshop)

Tao Zha

Federal Reserve Bank of Atlanta

October 2005

Table of contents

Table of contents (continued)

Table of contents (continued)

Table of contents (continued)

- Model II: SZW hyperinflation model for Latin American countries.
  - A hidden Markov-switching, adaptive expectations model.
  - The likelihood function.
  - Escapes, mean dynamics, rational expectations, and SCE.
  - Quantitative implications of our estimates for rational expectations and SCE versions.

References

- Thomas J. Sargent, 1999. *The Conquest of American Inflation*, Princeton Univeristy Press.
- John Geweke, 2005. *Contemporary Bayesian Econometrics and Statistics*, Wiley.
- Gelman, Carlin, Stern, and Rubin, 2004. *Bayesian Data Analysis* (second edition), Chapman & Hall/CRC.
- Sargent,Williams, and Zha, 2005. “Shocks and Government Beliefs: The Rise and Fall of American Inflation.”
- Sargent,Williams, and Zha, 2005. “Hyperinflations: expectations, escapes, and regime switches.”

General framework (rational expectations)

A rational expectations (RE) version of a class of dynamic stochastic general equilibrium(DSGE) models:

$$F_t \left(y_t, y_{t-1}, E_t y_{t+1}, E_{t-1} y_t, z_t, \epsilon_t \mid \phi \right) = 0, \tag{1}$$

where y_t is an $m \times 1$ vector formed by stacking all the variables in the model that have an expectational term, ϵ_t is an $n \times 1$ vector of independent random disturbances, ϕ is a vector of all model parameters, and z_t is a vector formed by stacking all current and lagged variables that have no expectational terms.

General framework (adaptive expectations)

A learning version of the RE model is to replace $E_t y_{t+1}$ by

$$E_t^{\mathsf{b}} y_{t+1} \equiv \beta_t, \tag{2}$$

where the superscript b stands for learning. Denote the *i*th element of β_t by β_{it}, which has the following general functional form f_i^{b}:

$$\beta_{it} = f_i^{\mathsf{b}} \left(y_t, y_{t-1}, x_{it}, x_{it-1} \mid \alpha_{t|t-1}, \alpha_{t-1|t-2}, P_{t|t-1}, P_{t-1|t-2}, V, \phi_o \right), \text{ for Kalman filter,} \tag{3}$$

$$\beta_{it} = f_i^{\mathsf{b}} \left(y_t, y_{t-1}, x_{it}, x_{it-1} \mid \alpha_{t|t-1}, \alpha_{t-1|t-2}, P_{t|t-1}, P_{t-1|t-2}, g, \phi_o \right), \text{ for constant gain,} \tag{4}$$

where x_{it}, a subset of current and lagged values of y_t and z_t, is a vector of right-hand-side observable variables (up to t) in the learning regression, $\alpha_{t|t-1}$ is a vector of beliefs based on information up to $t-1$, $P_{t|t-1}$ and V (or g) governs the dynamic evolution of $\alpha_{t+1|t}$, and ϕ_o is a vector of other model parameters. Without loss of generality, we let $m = 1$ (the dimension of y_t be one) and thus omit the subscript i in (3) and (4) for the rest of the lectures. Note that the vector x_t has the same length as $\alpha_{t|t-1}$.

Learning mechanisms

The learning mechanisms studied in the literature are based on the following regression form:

$$y_t = \alpha'_{t|t-1} x_t + \sigma w_t, \tag{5}$$

where w_t is an i.i.d. standard normal random variable.

Given $\alpha_{1|0}$ and $P_{1|0}$, a Kalman filter algorithm updates $\alpha_{t|t-1}$ for $t \geq 2$ with the following formula:

$$\alpha_{t|t-1} = \alpha_{t-1|t-2} + \frac{P_{t-1|t-2} x_{t-1} (y_{t-1} - x'_{t-1} \alpha_{t-1|t-2})}{\sigma^2 + x'_{t-1} P_{t-1|t-2} x_{t-1}}, \tag{6}$$

$$P_{t|t-1} = P_{t-1|t-2} - \frac{P_{t-1|t-2} (x_{t-1} x'_{t-1}) P_{t-1|t-2}}{\sigma^2 + x'_{t-1} P_{t-1|t-2} x_{t-1}} + V. \tag{7}$$

Learning mechanisms (continued)

Given $\alpha_{1|0}$ and $P_{1|0}$, a constant-gain least squares algorithm updates $\alpha_{t|t-1}$ for $t \geq 2$ with the following formula:

$$\alpha_{t|t-1} = \alpha_{t-1|t-2} + gP_{t-1|t-2}x_{t-1}(y_{t-1} - x'_{t-1}\alpha_{t-1|t-2}), \tag{8}$$

$$P^{-1}_{t|t-1} = P^{-1}_{t-1|t-2} + g\big(x_{t-1}x'_{t-1} - P^{-1}_{t-1|t-2}\big). \tag{9}$$

It can be shown from (6)-(9) that a constant-gain algorithm can be approximated by a Kalman filter with V proportional to $\sigma^2 E(x_t x'_t)^{-1}$. This approximation works particularly well for mean dynamics (which are described below).

Intuitively speaking, the *loose* relationships between the Kalman-filter and constant-gain algorithms can be summarized as follows.

- Large V corresponds to large g. Both imply that the past data are heavily discounted when beliefs are updated.
- A constant gain implies that V may be time-varying, proportional to $P_{t|t-1}$.
- The Kalman filter algorithm implies that the gain may be time-varying.

Mean dynamics – Kalman filter

It is difficult to obtain analytical asymptotic results for the dynamics implied by arbitrary V. For small V's, the beliefs drift at a slower rate and their evolution can be well approximated by the mean dynamics, which are defined as follows. Consider

$$\alpha_{t|t-1} = \alpha_{t-1|t-2} + \frac{(P_{t-1|t-2}/\epsilon)x_{t-1}(y_{t-1} - x'_{t-1}\alpha_{t-1|t-2})}{\sigma^2 + x'_{t-1}(P_{t-1|t-2}/\epsilon)x_{t-1}},$$

$$(P_{t|t-1}/\epsilon) = (P_{t-1|t-2}/\epsilon) - \frac{(P_{t-1|t-2}/\epsilon)(x_{t-1}x'_{t-1})(P_{t-1|t-2}/\epsilon)}{\sigma^2 + x'_{t-1}(P_{t-1|t-2}/\epsilon)x_{t-1}} + \epsilon^2 V.$$

As $\epsilon \longrightarrow 0$, the sequence of $\{\alpha_{t|t-1}, P_{t|t-1}/\epsilon\}$ generated by the above converges weakly to the solution of the following ordinary differential equations (ODEs):

$$\dot{\alpha} = PE\left[x_t(y_t - x'_t\alpha)\right], \tag{10}$$

$$\dot{P} = \sigma^{-2}V - PE(x_t x'_t)P. \tag{11}$$

The dynamics generated by (10) and (11) are called mean dynamics.

Mean dynamics – constant gain

For a constant-gain learning algorithm, the beliefs will drift at a slow rate when g is small. Consider

$$\alpha_{t|t-1} = \alpha_{t-1|t-2} + \epsilon g P_{t-1|t-2} x_{t-1} (y_{t-1} - x'_{t-1}\alpha_{t-1|t-2}),$$
$$P^{-1}_{t|t-1} = P^{-1}_{t-1|t-2} + \epsilon g \big(x_{t-1}x'_{t-1} - P^{-1}_{t-1|t-2}\big).$$

As $\epsilon \longrightarrow 0$, the sequence of $\{\alpha_{t|t-1}, P^{-1}_{t|t-1}\}$ generated by the above converges weakly to the solution of the following ordinary differential equations (ODEs):

$$\frac{d\alpha(t)}{dt} = PE\left[x_t(y_t - x'_t\alpha)\right], \tag{12}$$
$$\frac{dP^{-1}(t)}{dt} = E(x_t x'_t) - P^{-1}. \tag{13}$$

The above ODEs generate the mean dynamics.

Self-confirming equilibrium

A self-confirming equilibrium (SCE) is a vector of beliefs α_{SCE} that is consistent with what it observes and satisfies the population orthogonality and moment conditions implied by the ODEs (10) and (11) ((12) and (13)) governing the mean dynamics:

$$E\left(x_t(y_t - x_t'\alpha_{\text{SCE}})\right) = 0, \tag{14}$$

$$\sigma^{-2}V - P_{\text{SCE}}E(x_t x_t')P_{\text{SCE}} = 0, \text{ for Kalman filter,} \tag{15}$$

$$E(x_t x_t') - P_{\text{SCE}}^{-1} = 0, \text{ for constant gain.} \tag{16}$$

- If $V = \sigma^2 E x_t x_t'$, the SCE under the constant-gain learning is the same as the SCE under the Kalman filter learning, and the mean dynamics under both learning algorithms should be very close.
- For ϵ being not small (e.g., $\epsilon = 1$), the mean dynamics may not fully characterize the evolution of beliefs if V or g is relatively large. In this case, loosely speaking, we may get convergence to a nontrivial limit distribution of beliefs that are governed by both mean and *escape* dynamics.
- An SCE serves as a natural bridge between a RE equilibrium and an adaptive expectations equilibrium.

Maximum likelihood (ML) procedure – the block-wise algorithm

Let d_t, a subset of y_t and z_t, be an $n \times 1$ vector of observable variables corresponding to the fundamental shocks ϵ_t in (1), and s_t be a vector of hidden or observed exogenous variables in (1). Denote

$$D_t = \{d_0, \ldots, d_t\},\ S_t = \{s_0, \ldots, s_t\}.$$

Using a state-space form of (1), suppose that one can form the conditional likelihood function

$$p(d_t | D_{t-1}, S_t, \phi). \tag{17}$$

- If s_t is a vector of hidden exogenous variables whose distribution is assumed to be known, one can in principle integrate out S_T in (17) to get the overall likelihood $p(D_T|\phi)$.
- If s_t is a vector of observed exogenous variables, then

$$p(D_T | S_T, \phi) = \prod_{t=1}^{T} p(d_t | D_{t-1}, S_t, \phi).$$

The maximum likelihood estimate (MLE) of ϕ can be obtained by maximizing the overall likelihood with an efficient optimization routine.

ML procedure – the block-wise algorithm (continued)

- When the model parameters ϕ are high-dimentional, the overall likelihood tend to have long ridges and local peaks. This makes any optimization routine difficult to perform.
- In this case, we break ϕ into several blocks following the idea of Gibbs sampling (which will discussed later) and maximize the likelihood over one block of parameters while conditioning on the other blocks of parameters at the previous values.
- One iterates on this optimization block by block until the overall convergence criterion is satisfied.
- One particular block of parameters contains some key learning parameters that must be estimated to fit to the data (reverse-engineering estimation).
- These parameters are V, $P_{1|0}$, and g.

ML procedure – the block-wise algorithm (continued)

- It may be necessary to estimate the initial belief coefficients $\alpha_{1|0}$, depending on whether such estimation may create the over-fitting problem or whether it may destroy a reasonable economic interpretation of belief coefficients.
- In the SWZ US inflation model, $\alpha_{1|0}$ is fixed at the regression estimate obtained from the presample data so as to avoid the unduly influence by the updated beliefs in the sample.
- The parameter σ is in general not a free parameter and is normalized according to a certain rule (we will discuss this issue further in the SWZ US inflation paper).

ML procedure – posterior odds ratios

Without loss of generality, we omit the exogenous variables s_t for the rest of the lectures. Suppose we have two models; let ϕ_{m1} correspond to the $r_1 \times 1$ vector of parameters for Model 1 and ϕ_{m2} be the $r_2 \times 1$ vector for Model 2. The overall likelihood functions for both models are $p(D_T|\phi_{m1})$ and $p(D_T|\phi_{m2})$. Let the degrees of freedom for Model 2 be $df = r_2 - r_1$, and $\hat{\phi}_{m1}$ and $\hat{\phi}_{m2}$ be the MLEs respectively. By the Schwarz criterion (SC), the asymptotic posterior odds ratio of Model 1 to Model 2 in log value is

$$\text{log}p(D_T|\hat{\phi}_{m1}) - \Big(\text{log}p(D_T|\hat{\phi}_{m2}) - (df/2)\text{log}T\Big) .$$

Note that the log likelihood for Model 2 is adjusted by the degrees of freedom (relative to the baseline model, which is Model 1). Thus, as the number of the model's parameters increases, the model will be penalized in terms of its odds.

Markov-Chain Monte Carlo (MCMC) procedure – the Gibbs sampler

The likelihood function oftentimes is unbounded when it is integrated over the parameter space. This makes is impossible to make probability (or Bayesian) inferences. In general, the likelihood function is multiplied by a proper prior distribution $p(\phi)$, which gives the posterior probability density:

$$p(\phi|D_T) \propto p(D_T|\phi)p(\phi).$$

Under fairly general conditions, the posterior pdf is proper as the prior is proper.

The Gibbs sampler begins with breaking ϕ into, say, B blocks of parameters where each block $\phi_{(b)}$ for $b \in \{1, \ldots, B\}$ has a know distribution conditional on the other blocks. Given the initial values $\phi_{(1)}^{(0)}, \ldots, \phi_{(B)}^{(0)}$ (usually taken from the estimate at the posterior peak or from the neighborhood around the peak to ensure that the starting value does not have an extremely low probability), one can make the successive drawings:

$$\phi_{(b)}^{(t)} \sim p\left(\phi_{(b)}|\phi_{<(b)}^{(t)}, \phi_{>(b)}^{(t-1)}, D_T\right),\ t = 1, 2, \ldots,$$

where $\phi_{<(b)}$ denotes all the blocks $\phi_{(i)}$ for $i < b$ and $\phi_{>(b)}$ denotes all the blocks $\phi_{(i)}$ for $i > b$.

MCMC procedure – the Metropolis algorithm

There is a more general algorithm called Metropolis-Hastings. But since the Metropolis algorithm is often used for many problems in economics, we describe the Metropolis algorithm only.

Given the initial value $\phi^{(0)}$ (often taken from the estimate at the posterior peak or from the neighborhood around the peak), the Metropolis algorithm involves the four steps for $t = 1, 2, \ldots$:

(1) Given the value $\phi^{(t-1)}$, compute the proposal draw

$$\phi^{\text{prop}} = \phi^{(t-1)} + \xi_\phi,$$

where $\xi_\phi \sim N(\mathfrak{c}\,\tilde{\Sigma}_\phi)$ (although ξ_ϕ can have any distribution as long as the pdf value for ϕ^{prop} conditional on $\phi^{(t-1)}$ is the same as the pdf value for $\phi^{(t-1)}$ conditional on ϕ^{prop}.

MCMC procedure – the Metropolis algorithm (continued)

(1) To capture the correlation among the elements in ϕ, the covariance matrix $\tilde{\Sigma}_\phi$ may be computed to be the inverse of the second derivatives matrix of the log likelihood formed as

$$\left[\sum_{t=1}^{T} \hat{g}_t \hat{g}_t'\right]^{-1}$$

where

$$\hat{g}_t = \frac{\partial \, \text{log} p(d_t | D_{t-1}, \hat{\phi})}{\partial \, \hat{\phi}}.$$

The scale factor $\mathfrak{c}$ will be adjusted to keep the acceptance ratio optimal (around $25\% - 40\%$).

(2) Compute

$$r = \min\left\{\frac{p(\phi^{\text{prop}} | D_T)}{p(\phi^{(t-1)} | D_T)}, \; 1\right\}.$$

(3) Randomly draw ν from the uniform distribution $U(0,1)$.

(4) If $\nu <= r$, let $\phi^{(t)} = \phi^{\text{prop}}$ (acceptance); otherwise, keep $\phi^{(t)} = \phi^{(t-1)}$ (rejection).

MCMC procedure – Metropolis within Gibbs

In some applications (like the SWZ US inflation model), a conditional posterior density $p(\phi_{(b)}|\phi_{-(b)}$ within the Gibbs sampler is not of any standard form. Note that ($\phi_{-(b)}$ represent all the other blocks of parameters than $\phi_{(b)}$. In this case, the Metropolis algorithm can be applied to this block by setting $\phi_{(b)}^{(t)} = \phi_{(b)}^{\text{prop}}$ with the acceptance probability

$$r = \text{min}\left\{ \frac{p\left(\phi_{<(b)}^{(t)}, \phi_{(b)}^{\text{prop}}, \phi_{>(b)}^{(t-1)}|D_T\right)}{p\left(\phi_{<(b)}^{(t)}, \phi_{\geq(b)}^{(t-1)}|D_T\right)},\ 1 \right\}.$$

Marginal data density (MDD)

The MDD measures the model's fit to the data and forms a basis for calculating the posterior odds ratios. Denote the support of $p(\phi|D_T)$ by Θ_ϕ. The MDD's analytical solution is

$$p(D_T) = \int_{\Theta_\phi} p(D_T \mid \phi) p(\phi) \, d\phi$$

There are a number of methods for approximating the above object numerically.

- One generic approach is the MHM procedure (Gelfand and Dey 1994; Geweke 1999).
- Another one is to utilize Gibbs sampling (Chib 1995).

The latter approach is in general very reliable and accurate if the Gibbs sampler exists, because each conditional posterior density can be evaluated in closed form.

MDD – the modified harmonic means (MHM) procedure

Let $g(\phi)$ be a *weighting* function that must be a pdf (*not* kernel) whose support is contained in Θ_p. The MHM method is based on the observation that

$$p(D_T)^{-1} = \int_{\Theta_\phi} \frac{g(\phi)}{p(D_T \mid \phi)p(\phi)}\, p(\phi \mid D_T)d\phi. \tag{18}$$

A numerical evaluation of the integral on the right hand side of (18) can be done through the Monte Carlo integration

$$\hat{p}(D_T)^{-1} = \sum_{i=1}^{N} \frac{g(\phi^{(i)})}{p(D_T \mid \phi^{(i)})p(\phi^{(i)})}, \tag{19}$$

where $\phi^{(i)}$ is the i^{th} draw of ϕ from the posterior distribution $p(\phi \mid D_T)$.

MDD – using Gibbs sampling

The MDD can be evaluated using the following identity

$$p(D_T) = \frac{p\left(D_T \mid \phi^*\right) p\left(\phi^*\right)}{p\left(\phi^* \mid D_T\right)},$$

for any $\phi^* \in \Theta_\phi$. Typically, $p\left(D_T \mid \phi^*\right)$ and $p\left(\phi^*\right)$ can be evaluated in closed form but $p\left(\phi^* \mid D_T\right)$ cannot. To evaluate $p\left(\phi^* \mid D_T\right)$ accurately, we decompose this term according to the Gibbs sampler:

$$p\left(\phi^* \mid D_T\right) = p\left(\phi^*_{(1)} \mid D_T\right) p\left(\phi^*_{(2)} \mid \phi^*_{(1)}, D_T\right) \ldots p\left(\phi^*_{(B)} \mid \phi^*_{<(B)}, D_T\right).$$

The first term on the right hand side can be approximated well from the output of the original Gibbs sampler $\phi^{(i)}_{(b)}$:

$$N^{-1} \sum_{i=1}^{N} p\left(\phi^*_{(1)} \mid \phi^{(i)}_{>(1)}, D_T\right) \xrightarrow{a.s.} p\left(\phi^*_{(1)} \mid D_T\right).$$

MDD – using Gibbs sampling (continued)

Because the conditional density $p\left(\phi^*_{(1)} \mid \phi^{(i)}_{>(1)}, D_T\right)$ can be evaluated in closed form, the approximation will be accurate with enough Gibbs draws.

The last term $p\left(\phi^*_{(B)} \mid \phi^*_{<(B)}, D_T\right)$ can be evaluated in closed form so that no approximation is needed.

To approximate $p\left(\phi^*_{(b)} \mid \phi^*_{<(b)}, D_T\right)$, we use the Gibbs sampler to produce draws $\phi^{(i)\,(b-1)}_{>(b)}$ with the first $b-1$ blocks fixed at the $*$ values:

$$N^{-1} \sum_{i=1}^{N} p\left(\phi^*_{(b)} \mid \phi^*_{<(b)}, \phi^{(i)\,(b-1)}_{>(b)}, D_T\right) \xrightarrow{a.s.} p\left(\phi^*_{(b)} \mid \phi^*_{<(b)}, D_T\right).$$

Again all the conditional posterior densities can be evaluated in closed form. This algorithm is efficient if ϕ^* is chosen near the posterior mode. By varying the value of ϕ^*, one can

- obtain the standard error for the MDD,
- use it as the basis to check the convergence of the Gibbs sampler.

VARs as a benchmark for model comparison

- VARs have been known to fit to the data very well and thus have often used as a benchmark for comparing different DSGE models.
- *Restricted* VARs have been used to give each structural equation more economic interpretations (Rudebusch and Svensson 1999, 2002).

Is there any efficient method to compute the MDD for a restricted VAR? Waggnoer and Zha (2003) develop a Gibbs sampling algorithm for a class of identified VARs with linear restrictions on both contemporaneous and lagged coefficients. The models of Rudebusch and Svensson (1999, 2002) are special cases of these restricted VARs.

- With the Gibbs sampler for restricted VARs, one can use the Chib procedure discussed above to obtain a very accurate evaluation of the MDD.
- For example, it takes less than 1 minute to obtain a reliable estimate of the MDD for a large VAR with 13 lags and 10 variables.
- The program swz_mardd.m computes the MDD for a restricted VAR. The file readme_swz.prn gives detailed instructions of how to use the code.
- For quarterly data, the prior is typically set as mu(1) = 1; mu(2) = 0.5 (or 0.2); mu(3) = 0.1; mu(4) = 1.0; mu(5) = 1.0; mu(6) = 1.0.
- For monthly data, it is typically set as mu(1) = 0.6; mu(2) = 0.1; mu(3) = 0.1; mu(4) = 1; mu(5) = 5.0; mu(6) = 5.0.

The SWZ US inflation model

The SWZ model takes a special form of (1) where $y_t = \pi_t$, $z_t = u_t$, and $\epsilon_t = [w_{1t}\ w_{2t}]'$:

$$u_t - u^* = \theta_0(\pi_t - E_{t-1}\pi_t) + \theta_1(\pi_{t-1} - E_{t-2}\pi_{t-1}) + \tau_1(u_{t-1} - u^*) + \sigma_1 w_{1t}, \quad (20)$$

$$\pi_t = x_{t-1} + \sigma_2 w_{2t}. \quad (21)$$

- Equation (20) is an expectations-augmented Phillips curve in which systematic monetary policy has neither short-run nor long-run effects on unemployment.
- Equation (20) embodies a stronger form of ‘policy irrelevance’ than do many of today’s New Keynesian Phillips curves.
- Ignore the nonneutralities present in those models and aim to reverse engineer a set of government beliefs that can explain the low frequency swings in U.S. data.
- Insist that the true DGM have the strong policy irrelevance of the Lucas supply function.

Government's optimization problem

The policy decision x_{t-1} solves the "Phelps problem":

$$\min_{x_{t-1}} \hat{E} \sum_{t=1}^{\infty} \delta^t ((\pi_t - \pi^*)^2 + \lambda(u_t - u^{**})^2) \tag{22}$$

subject to (21) and

$$u_t = \hat{\alpha}'_{t|t-1} \Phi_t + \sigma w_t. \tag{23}$$

The beliefs $\hat{\alpha}_{t|t-1}$ is updated in the same way as (6) and (7), where we replace $\alpha_{t|t-1}$ by $\hat{\alpha}_{t|t-1}$, x_t by Φ_t, and y_t by u_t.

Solving this government problem leads to the solution characterized by (3). Specifically, we will have

$$x_{t-1} = f^{\mathrm{b}}(u_{t-1}, \Phi_{t-1} | \hat{\alpha}_{t|t-1}, P_{t|t-1}, V, \sigma, \delta, \lambda, \pi^*, u^{**}). \tag{24}$$

Omitting the subscript $t|t-1$ for the α's, we consider

$$u_t = \hat{\alpha}_0 \pi_0 + \hat{\alpha}_1 \pi_{t-1} + \hat{\alpha}_2 u_{t-1} + \hat{\alpha}_3 \pi_{t-1} + \hat{\alpha}_4 u_{t-2} + \hat{\alpha}_5 + \sigma w_t. \tag{25}$$

Government's optimization problem (Euler equations)

Let L_{1t} be the Lagrangian multiplier for (21) and L_{2t} for (25). The first-order conditions w.r.t. u_t, π_t, and x_{t-1} are

$$2\lambda(u_t - u^{**}) = L_{2t} - \delta\hat{\alpha}_2 E_t L_{2t+1} - \delta^2\hat{\alpha}_4 E_t L_{2t+2}, \tag{26}$$

$$2(\pi_t - \pi^*) = \hat{\alpha}_0 L_{2t} - \delta\hat{\alpha}_1 E_t L_{2t+1} - \delta^2\hat{\alpha}_3 E_t L_{2t+2} + L_{1t}, \tag{27}$$

$$E_{t-1} L_{1t} = 0. \tag{28}$$

Two dummy equations are

$$\pi_{t-1} = \pi_{t-1}, \tag{29}$$

$$u_{t-1} = u_{t-1}, . \tag{30}$$

Three expectational errors are ξ_{1t}, ξ_{2t}, and ξ_{3t} defined as

$$L_{1t} = E_{t-1}L_{1t} + \xi_{1t} = \xi_{1t}, \tag{31}$$

$$L_{2t} = E_{t-1}L_{2t} + \xi_{2t}, \tag{32}$$

$$E_t L_{2t+1} = E_{t-1}L_{2t+1} + \xi_{3t}. \tag{33}$$

Sims's gensys form

To put the government's optimal problem in the gensys form (a linear form of (1)), let y_t^{G} be a vector of 8 variables:

$$y_t^{\mathsf{G}} = [x_t \ \pi_t \ \pi_{t-1} \ u_t \ u_{t-1} \ L_{2t} \ E_t L_{2t+1} \ E_t L_{2t+2}]'.$$

The canonical form for a rational expectations model is

$$\Gamma_0 y_t^{\mathsf{G}} = \Gamma_1 y_{t-1}^{\mathsf{G}} + c + \Psi \epsilon_t + \Pi \eta_t, \tag{34}$$

where ϵ_t is a vector of exogenous variables and η_t is a vector of endogenous disturbances satisfying $E_t \eta_{t+1} = 0$.

- There are 8 equations for this government problem.
- They are (21), (25), (26), (27) combined with (31), (29), (30), (32), and (33).
- We now put them in the matrices Γ_0, Γ_1, c, Ψ, and Π.

Canonical form for the government problem

$$\Gamma_0 = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -\hat{\alpha}_0 & -\hat{\alpha}_1 & 1 & -\hat{\alpha}_2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2\lambda & 0 & -1 & \delta\hat{\alpha}_2 & \delta^2\hat{\alpha}_4 \\ 0 & 2 & 0 & 0 & 0 & \hat{\alpha}_0 & \delta\hat{\alpha}_1 & \delta^2\hat{\alpha}_3 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

$$\Gamma_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \hat{\alpha}_3 & 0 & \hat{\alpha}_4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Canonical form for the government problem (continued)

$$C = \begin{bmatrix} 0 \\ \hat{\alpha}_5 \\ 2\lambda u^{**} \\ 2\pi^* \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad \Psi = \begin{bmatrix} \sigma_2 & 0 \\ 0 & \sigma \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \quad \Pi = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \epsilon_t = \begin{bmatrix} w_{2t} \\ w_t \end{bmatrix} \quad \eta_t = \begin{bmatrix} \xi_{1t} \\ \xi_{2t} \\ \xi_{3t} \end{bmatrix}$$

Solving this optimization problem will lead to the solution of the form (24).

Ramsey equilibrium

As a benchmark, suppose that the government has full knowledge of the economy. In this case, there is no model misspecification on the part of the government, and the government solves out its optimal policy subject to (20) instead of (23).

This problem can be solved directly with Dynare's OLR procedure – using the Dynare module sz1. The optimal inflation policy for this case is always equal to π^*.

Nash equilibrium and SCE

The Nash equilibrium inflation rate is:

$$\pi^{\text{Nash}} = \pi^* - \lambda(u^* - u^{**}) \left[(1 + \delta\tau_1)\theta_0 + \delta\theta_1 \right]. \tag{35}$$

The larger are $u^* - u^{**}$, θ_0, and θ_1 in absolute value, the higher is the Nash inflation rate compared to the Ramsey rate π^*.

- The mean inflation rate at the SCE agrees with the Nash inflation rate.
- To obtain the SCE, one can numerically solve out the ODEs (10) and (11).
- Setting the government's beliefs at the SCE, the Dynare module sz2 can be used to simulate the dynamics of inflation, which may escape from the SCE.

Learning equilibrium

- The government has a misspecified model but updates its beliefs according to the Kalman filter algorithm.
- At each time t, the government's beliefs are updated and its optimal policy is resolved again according to the canonical form described above.
- The Dynare module sz3 solves this problem and simulates the path of inflation.

Role of the government's learning from misspecified models

- In this model, inflation converges much faster to the SCE under Kalman filtering learning than under RLS. In effect, the Kalman filter learning rule with drifting coefficients seems to discount the past data more rapidly than the constant gain RLS learning rule.
- Consider ϵV. As the government's prior belief parameter $\epsilon \longrightarrow 0$ (at the limit there is no time variation in the parameters), inflation converges to the self-confirming equilibrium (SCE) and the mean escape time becomes arbitrarily long.
- As the government's prior belief parameter $\sigma \longrightarrow 0$ (in the limit, there is no variation in the government's regression error or arbitrarily large time variation in the drifting parameters), large escapes from an SCE can happen arbitrarily often and nonconvergence is possible.
- The covariance matrix V in the government's prior belief about the volatility of the drifting parameters affects the speed of escape. The covariance matrix V combined with the prior belief parameter ϵ, affects the speed of convergence to the SCE from a low inflation level.

Methodology and MCMC algorithm

Group all other free structural parameters as

$$\phi = \{v^*, \theta_0, \theta_1, \tau_1, \zeta_1, \zeta_2, u(C_P), u(C_V)\},$$

where $v^* = u^*(1 - \tau_1)$, C_P and C_V are upper triangular such that $P_{1|0} = C_P' C_P$ and $V = C_V' C_V$, and $\zeta_1 = 1/\sigma_1^2$ and $\zeta_2 = 1/\sigma_2^2$ represent the precisions of the corresponding innovations. The notation $u(C_P)$ or $u(C_V)$ means that only the upper triangular part of C_P or C_V are among the free parameters.

The structural parameter $\zeta = 1/\sigma^2$ is not free. It is clear from (6) and (7) that if we scale V and $P_{1|0}$ by κ and ζ by $1/\kappa$, the beliefs remain the same and so does the likelihood (implying the model is unidentified). Following Sargent and Williams, we impose the restriction $\zeta = \zeta_1$.

To take into account parameter uncertainty, we employ the Bayesian method and develop a Monte Carlo Markov Chain (MCMC) algorithm that breaks ϕ into three separate blocks: θ, $\{\zeta_1, \zeta_2\}$, and φ where

$$\theta = [v^* \ \theta_0 \ \theta_1 \ \tau_1]',$$

and $\varphi = \{u(C_P), u(C_V)\}$.

Methodology and MCMC algorithm (continued)

The prior pdf of ϕ can be factored as:

$$p(\phi) = p(\theta)\, p(\varphi)\, p(\zeta_1, \zeta_2).$$

The likelihood function is:

$$\mathcal{L}\left(\mathcal{I}_T | \phi\right) = \frac{\zeta_1^{T/2} \zeta_2^{T/2}}{(2\pi)^{T/2}} \,\text{exp} \left\{ -\frac{1}{2} \sum_{t=1}^{T} \left[\zeta_1 z_{1t}^2 + \zeta_2 z_{2t}^2 \right] \right\}, \tag{36}$$

where z_{1t} and z_{2t} are the functions of θ and φ:

$$z_{1t} = u_t - u^* - \theta_0(\pi_t - x_{t-1}) - \theta_1(\pi_{t-1} - x_{t-2}) - \tau_1(u_{t-1} - u^*),$$
$$z_{2t} = \pi_t - x_{t-1},$$

where the optimal decision rule depends on φ.

Methodology and MCMC algorithm (continued)

The posterior pdf of ϕ is proportional to the product of the likelihood (36) and the prior $p(\phi)$:

$$p(\phi|\mathcal{I}_T) \propto \mathcal{L}\left(\mathcal{I}_T|\phi\right)\, p(\phi). \tag{37}$$

The posterior distribution of ϕ can be simulated by alternately sampling from the conditional posterior distributions (Metropolis with Gibbs):

$$p(\theta \mid \mathcal{I}_T, \zeta_1, \zeta_2, \varphi), \text{ Normal}$$

$$p(\zeta_1, \zeta_2 \mid \mathcal{I}_T, \theta, \varphi), \text{ Gamma}$$

$$p(\varphi \mid \mathcal{I}_T, \theta, \zeta_1, \zeta_2), \text{ Unknown pdf – using Metropolis.}$$

In estimation, we set $\delta = 0.9936, \lambda = 1, \pi^* = 2$, and $u^{**} = 1$. We set the initial belief $\hat{\alpha}_{1|0}$ at the regression estimate obtained from the presample data from January 1948 to December 1959.

Reverse engineering estimation

Log value of posterior kernel at its peak: 564.92

Estimates of coefficients in true Phillips curve and inflation process
with 68% and 90% probability intervals in parentheses

$u^* : \; 6.1104 \quad (5.2500, 7.1579) \quad (4.2238, 9.0586)$

$\theta_0 : \; -0.0008 \quad (-0.0237, 0.0475) \quad (-0.0458, 0.0719)$

$\theta_1 : \; -0.0122 \quad (-0.0375, 0.0297) \quad (-0.0589, 0.0526)$

$\tau_1 : \; 0.9892 \quad (0.9852, 0.9960) \quad (0.9817, 0.9996)$

$\zeta_1 : \; 35.6538 \quad (28.7565, 32.4947) \quad (27.6017, 33.7890)$

$\zeta_2 : \; 18.97671 \quad (15.6565, 18.2557) \quad (14.7008, 19.1196)$

Estimate of $P_{1|0}$:

10.8705	14.3236	2.2518	-25.4037	-0.9279	-10.1548
14.3236	19.3721	2.9624	-33.9832	-1.1883	-13.5923
2.2518	2.9624	0.4690	-5.2629	-0.1928	-2.1050
-25.4037	-33.9832	-5.2629	59.8997	2.1339	23.9551
-0.9279	-1.1883	-0.1928	2.1339	0.0816	0.8526
-10.1548	-13.5923	-2.1050	23.9551	0.8526	9.5810

Estimate of V:

8.2323	-7.7781	0.9208	4.9782	-0.8136	-41.414
-7.7781	8.1400	0.0303	-5.089	1.9353	68.591
0.9208	0.0303	2.9854	0.1187	3.7012	72.067
4.9782	-5.089	0.1187	3.2032	-1.0548	-39.963
-0.8136	1.9353	3.7012	-1.0548	5.1362	100.6400
-41.414	68.591	72.067	-39.963	100.6400	2588.3000

Reverse engineering estimation (continued)

Log likelihood and log posterior odds ratio

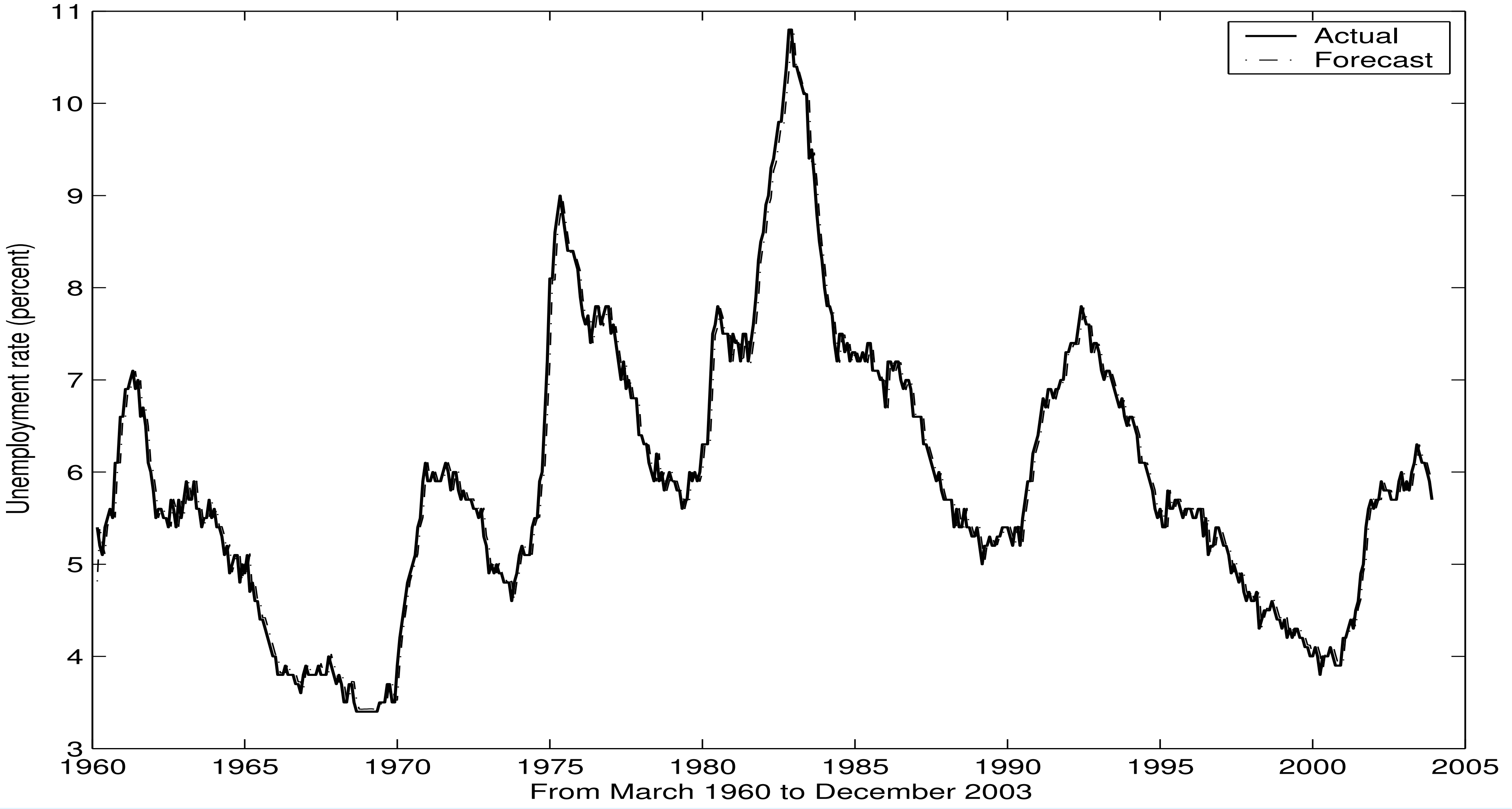
	SWZ model	BVAR(13)	log odds favoring SWZ
SC	564.92	309.37	255.55
MDD	424.75	244.65	180.10

Reverse engineering estimation (continued)

Log likelihood and log posterior odds ratio

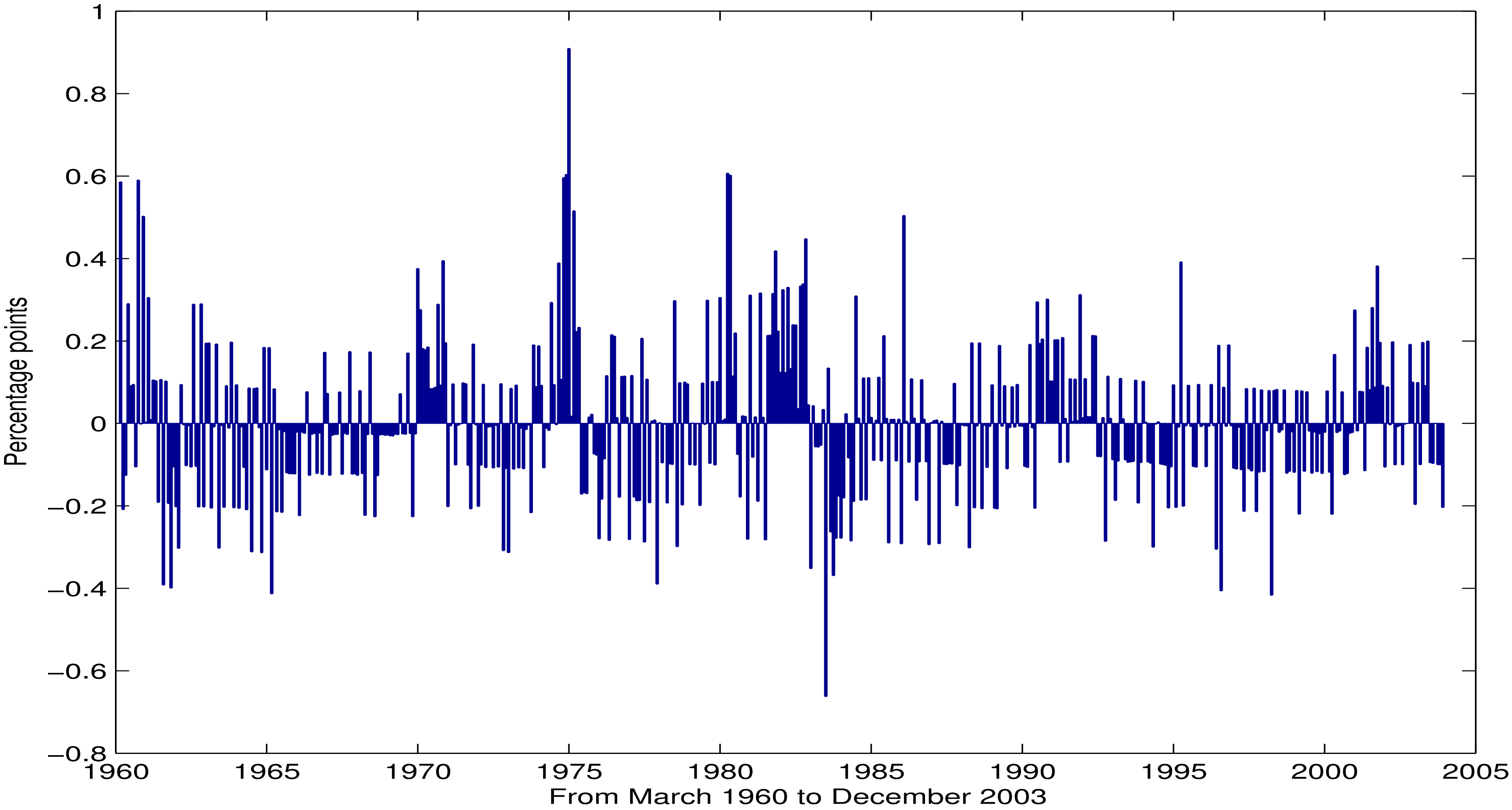
	SWZ model	BVAR(13)	log odds favoring SWZ
SC	564.92	309.37	255.55
MDD	424.75	244.65	180.10

Reverse engineering estimation (continued)



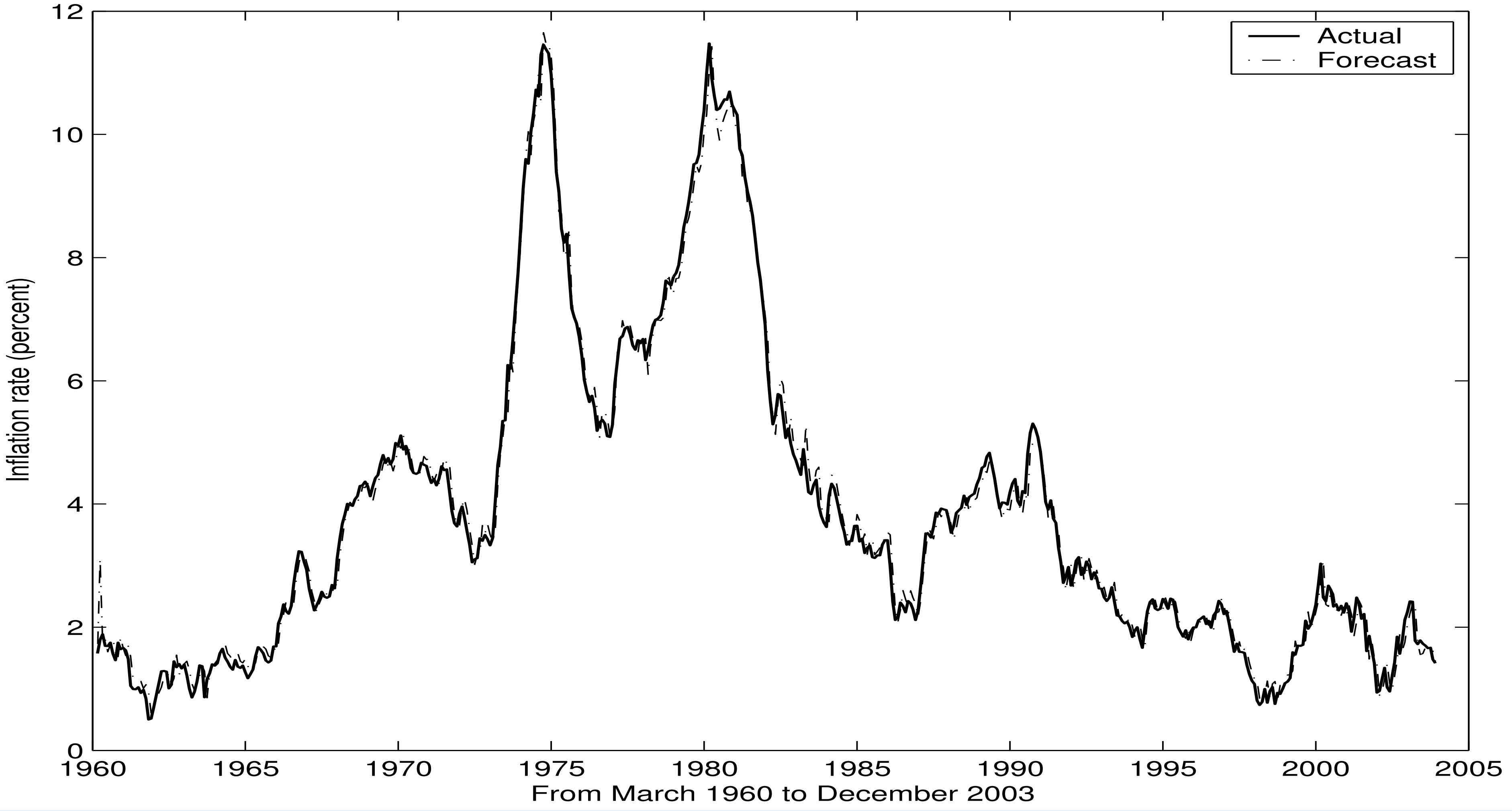
Unemployment rate: actual vs one-step forecast.

Reverse engineering estimation (continued)



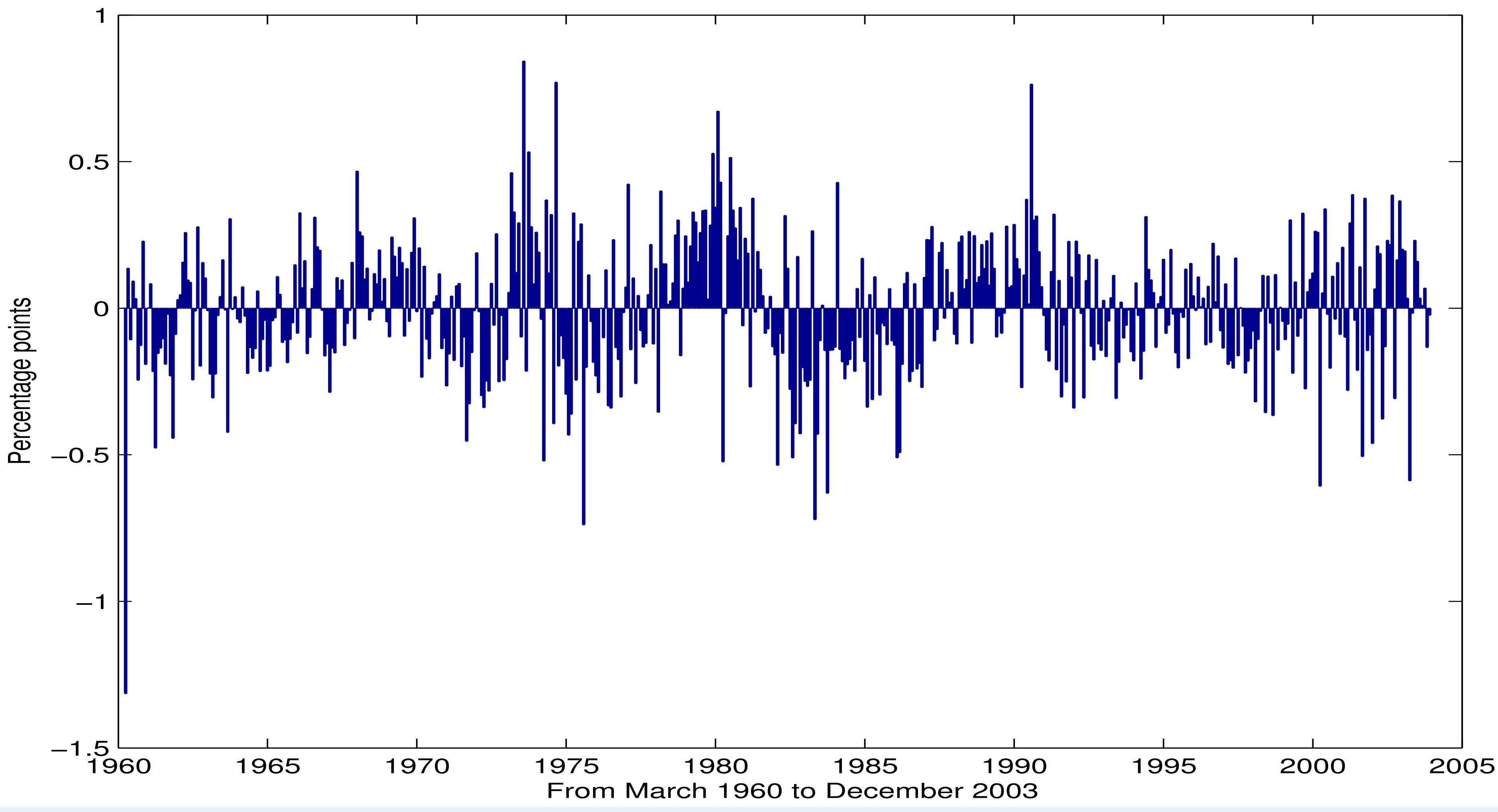
Differences between actual values and one-step forecasts of unemployment.

Reverse engineering estimation (continued)



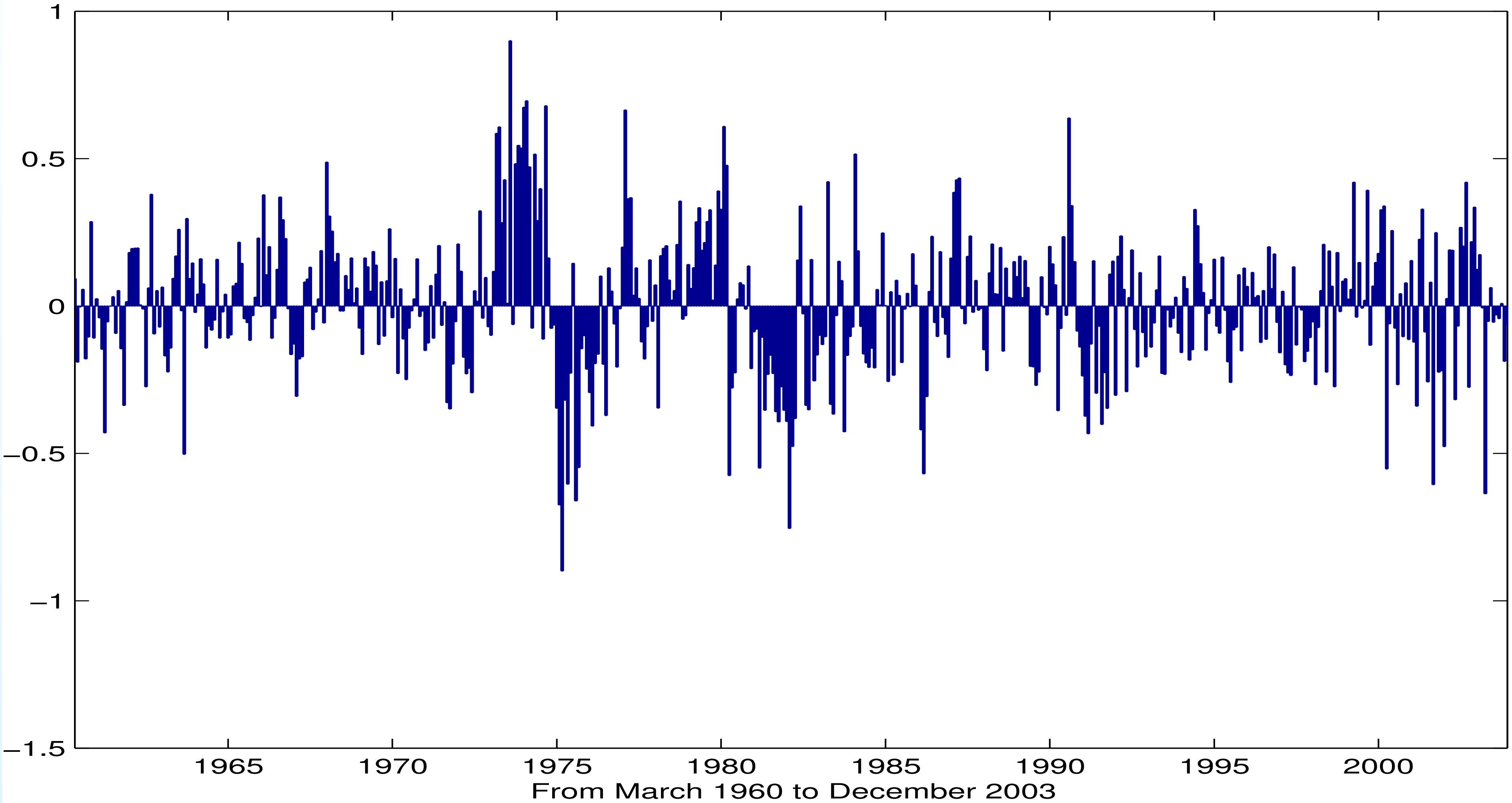
Inflation: actual vs one-step forecast.

Reverse engineering estimation (continued)



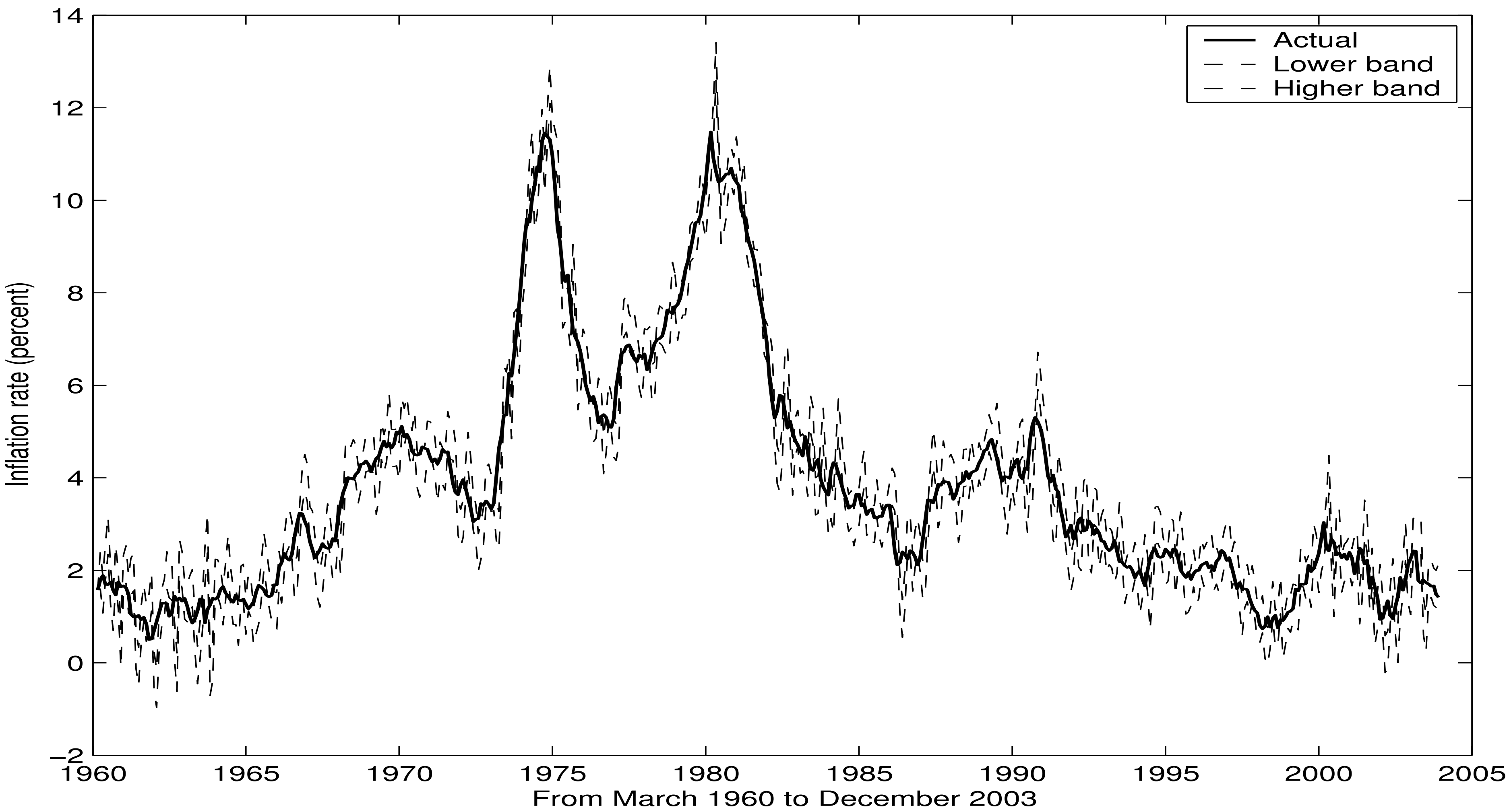
Differences between actual values and one-step forecasts of inflation.

Reverse engineering estimation (continued)



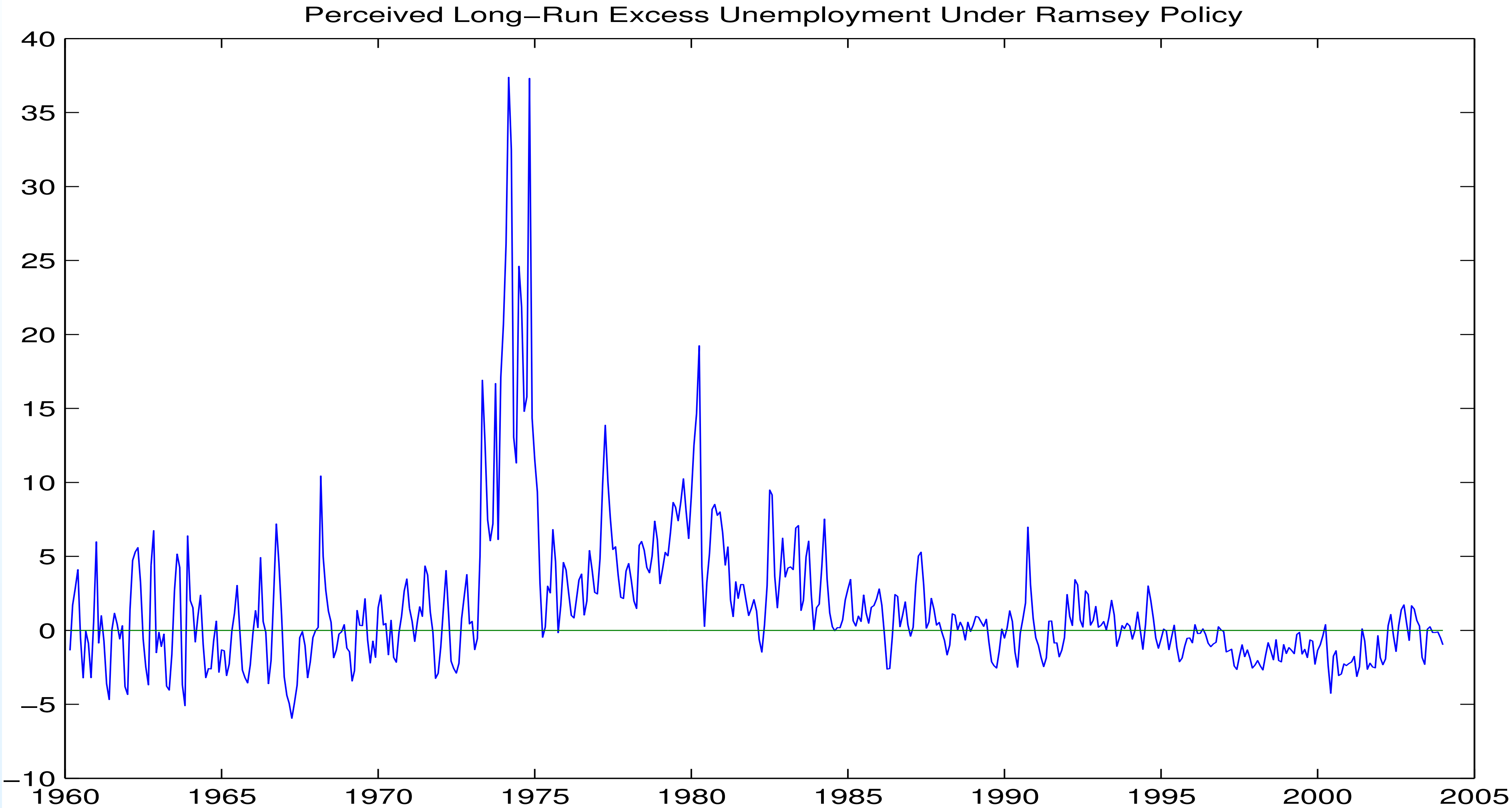
Random walk: differences between actual values and one-step forecasts of inflation.

Reverse engineering estimation (continued)



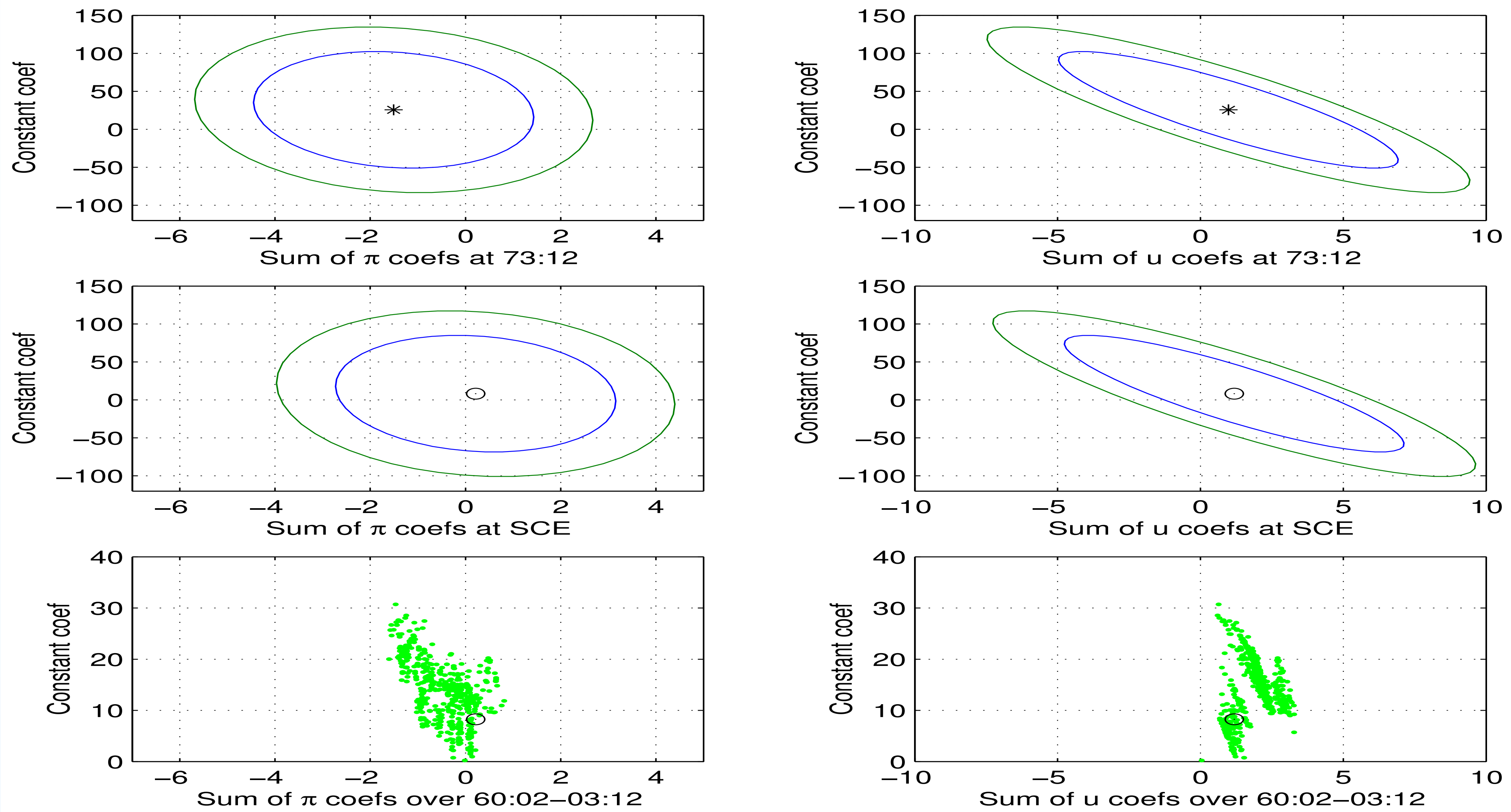
Inflation: actual vs one-step forecast with 90% error bands.

Shocks and beliefs



Perceived sacrifice ratios.

Shocks and beliefs (continued)

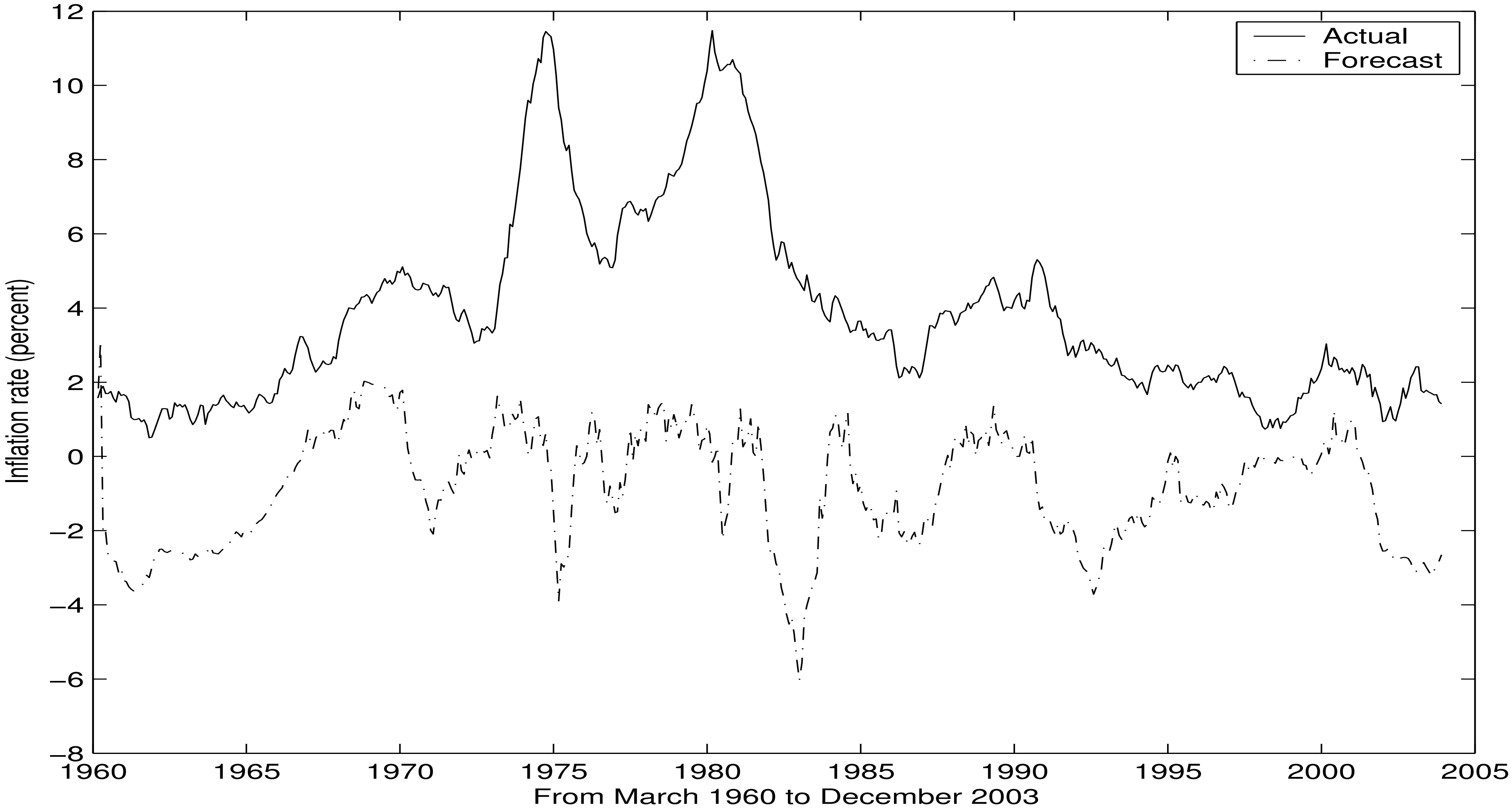


68% and 90% probability ellipses about key parameters in the government's Phillips curve.

Importance of cross-equation restrictions

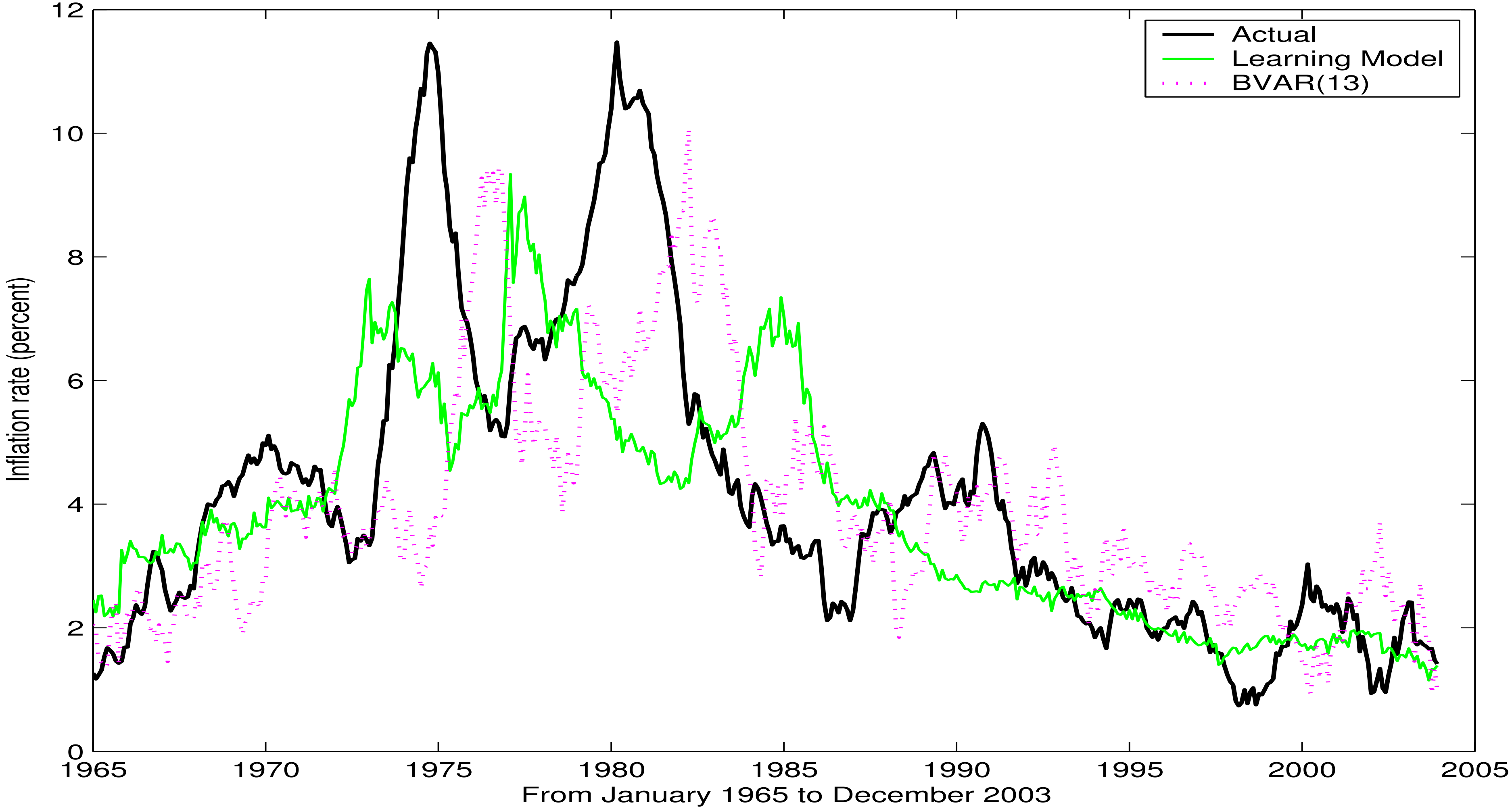
- The FOMC's preoccupation with the recent data (readings from the FOMC's transcripts).
- Large and consequential changes in the FRB/US model at the Federal Reserve Board from July 1996 and November 2003 (Tetlow and Ironside 2005).

Importance of cross-equation restrictions (continued)



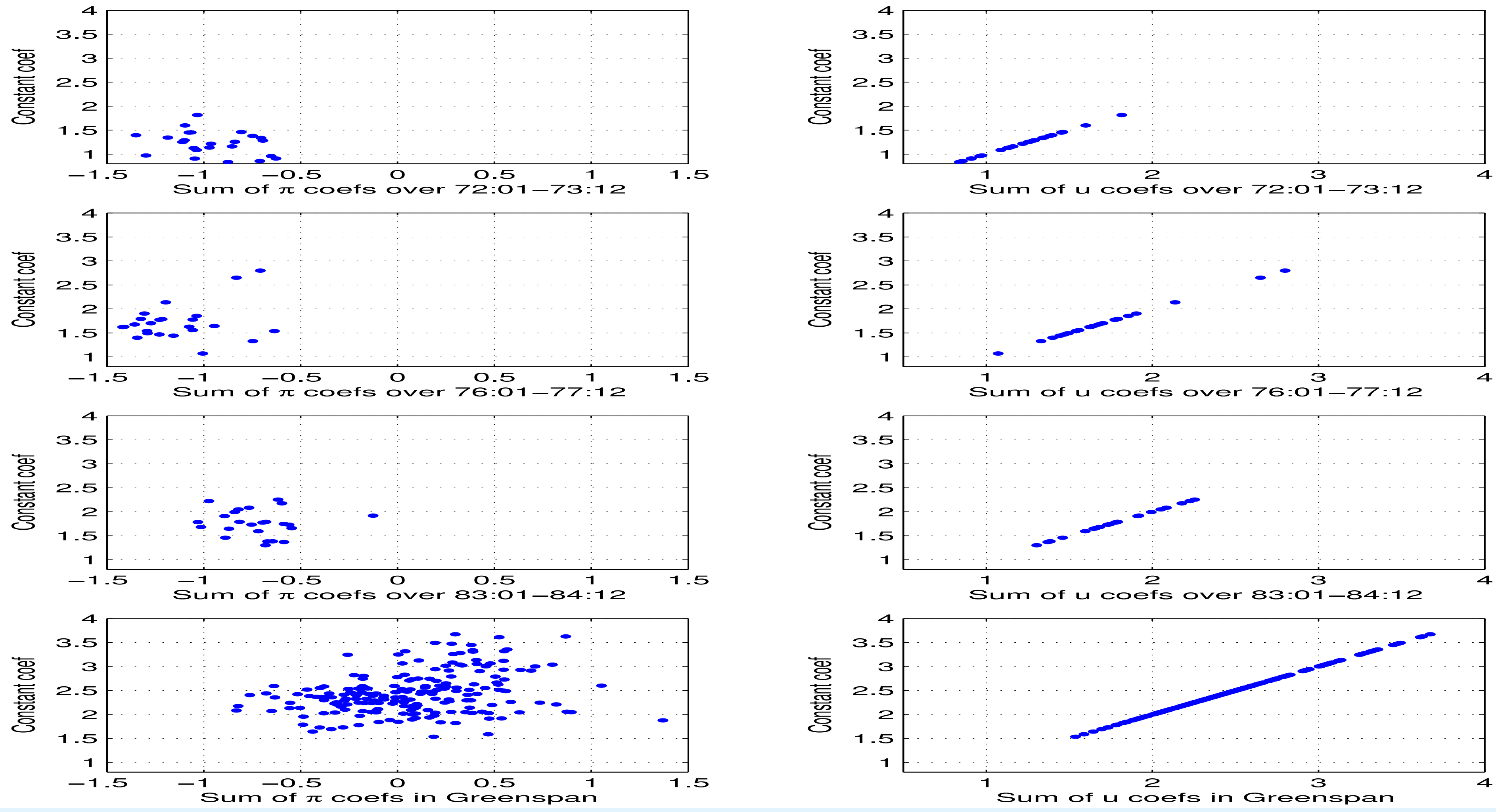
Actual inflation and one-step prediction from the benchmark model in which V is estimated without imposing the cross-equation restrictions

Importance of cross-equation restrictions (continued)



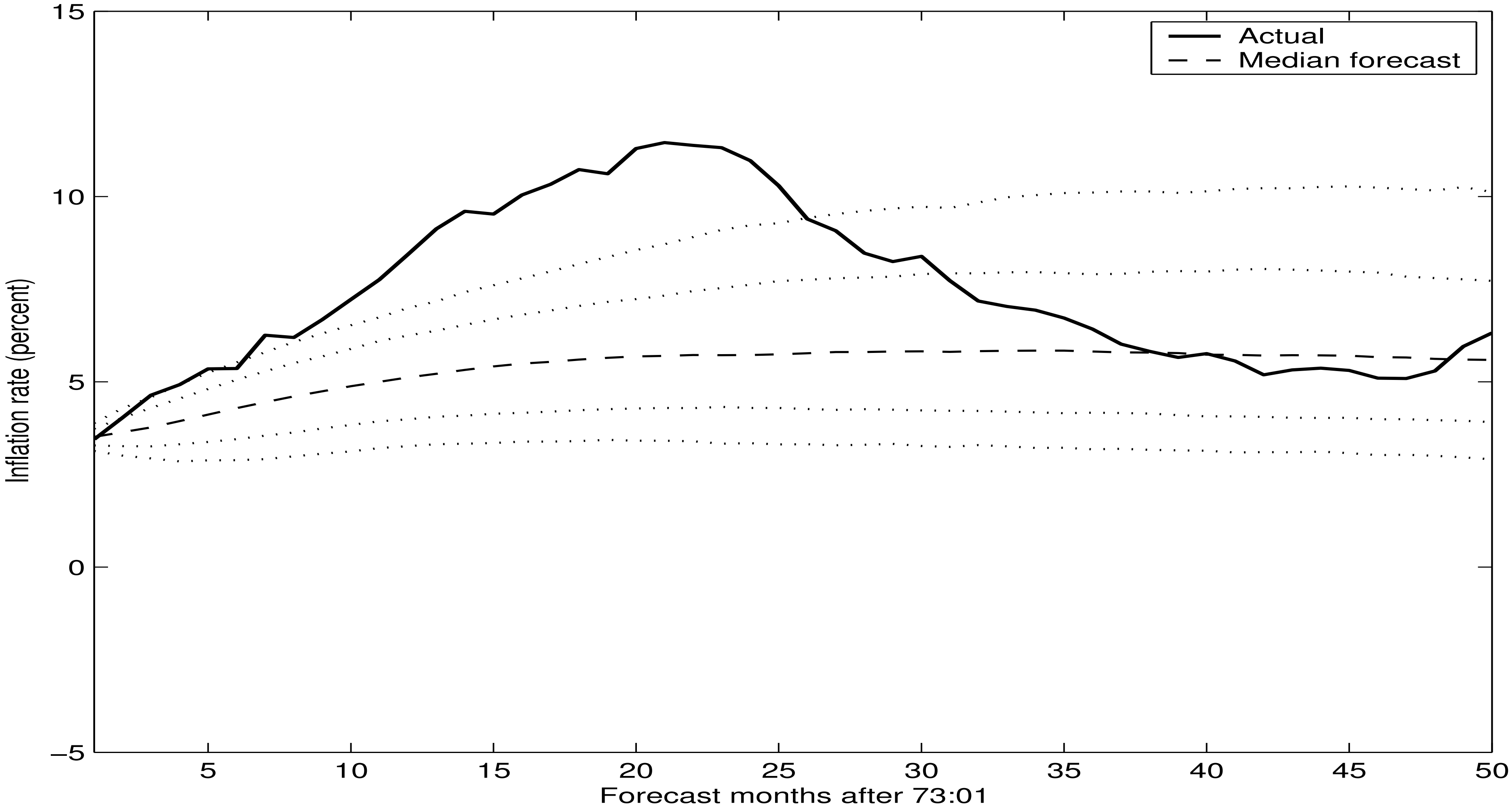
Two-year-ahead inflation forecasts: learning model versus BVAR(13)

Importance of cross-equation restrictions (continued)



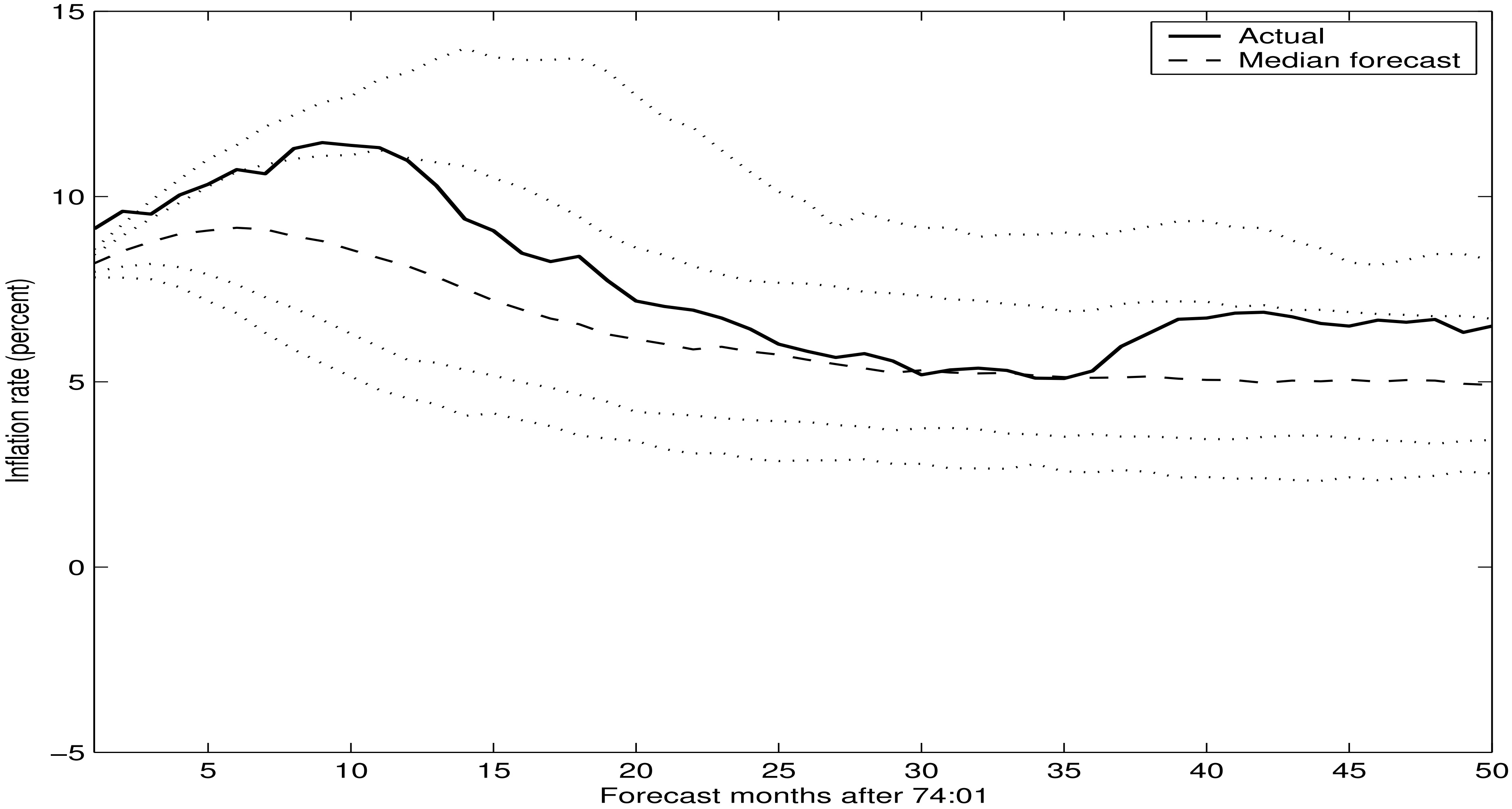
Estimates of key step-on-gas parameters over the three predicted run-up periods and over the Greenspan era

Two peaks and an enduring decline



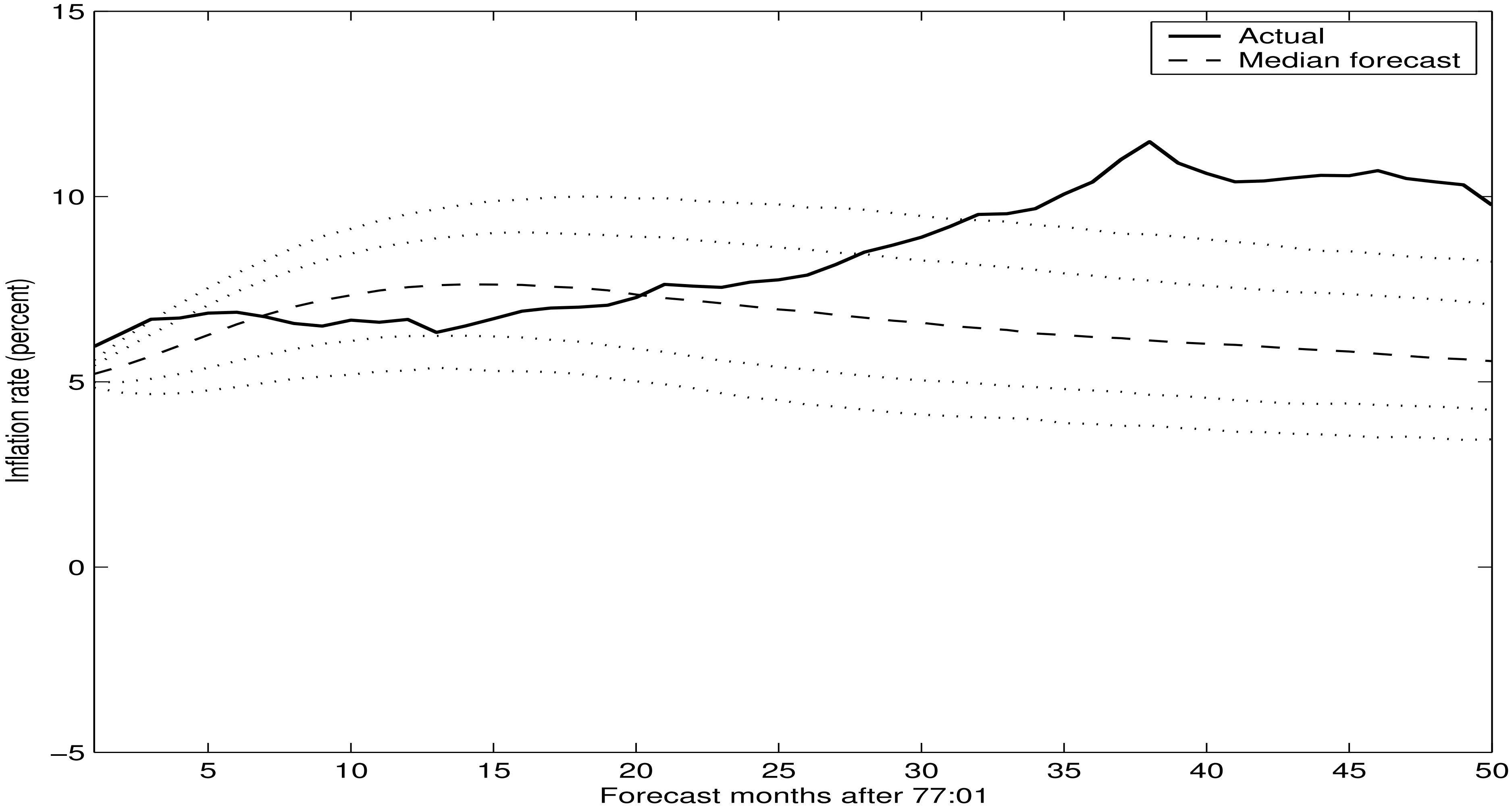
Dynamic forecasts with 68% and 90% error bands, using initial estimated condition at 73:01

Two peaks and an enduring decline (continued)



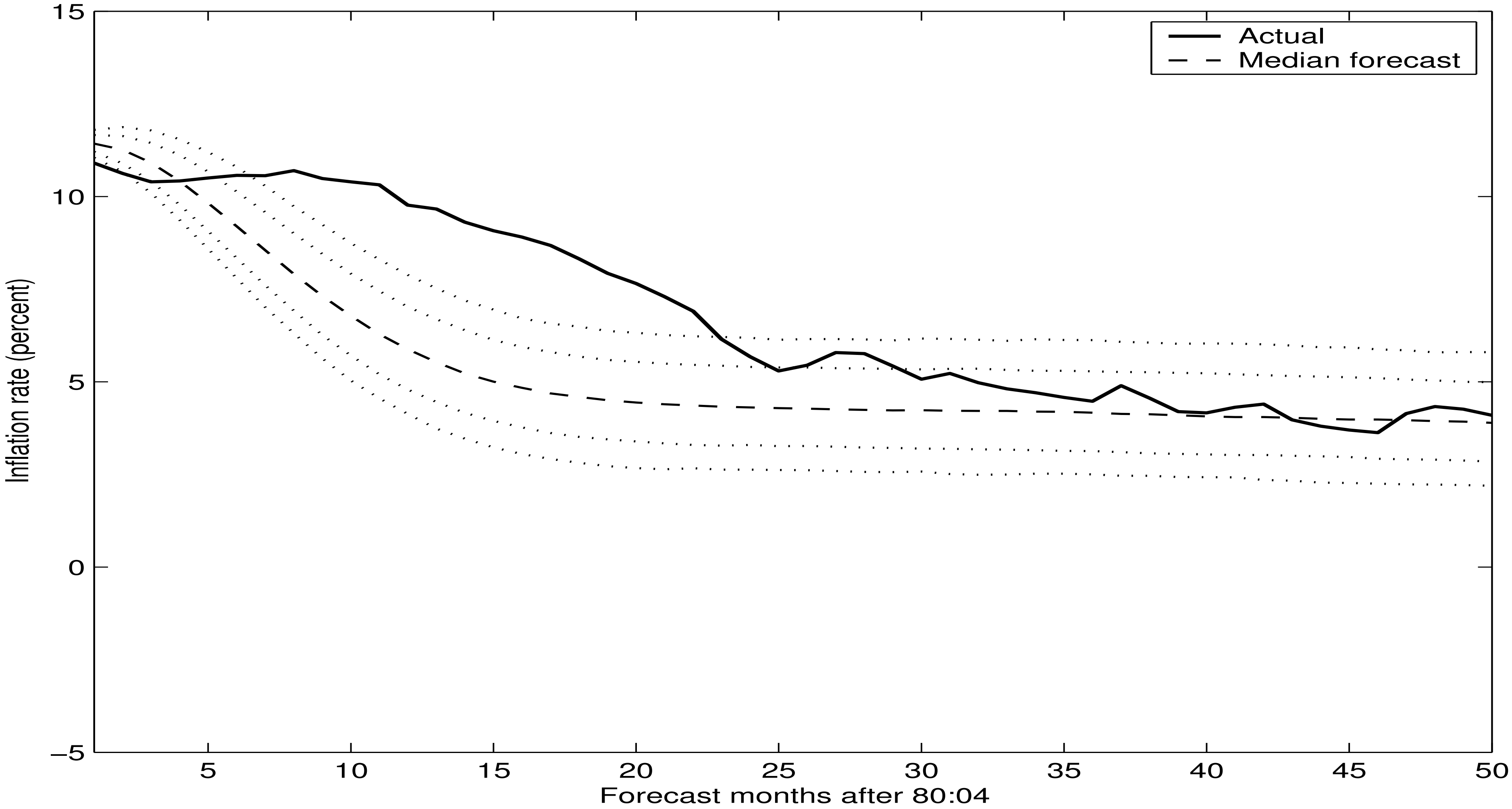
Dynamic forecasts with 68% and 90% error bands, using initial estimated condition at 74:01

Two peaks and an enduring decline (continued)



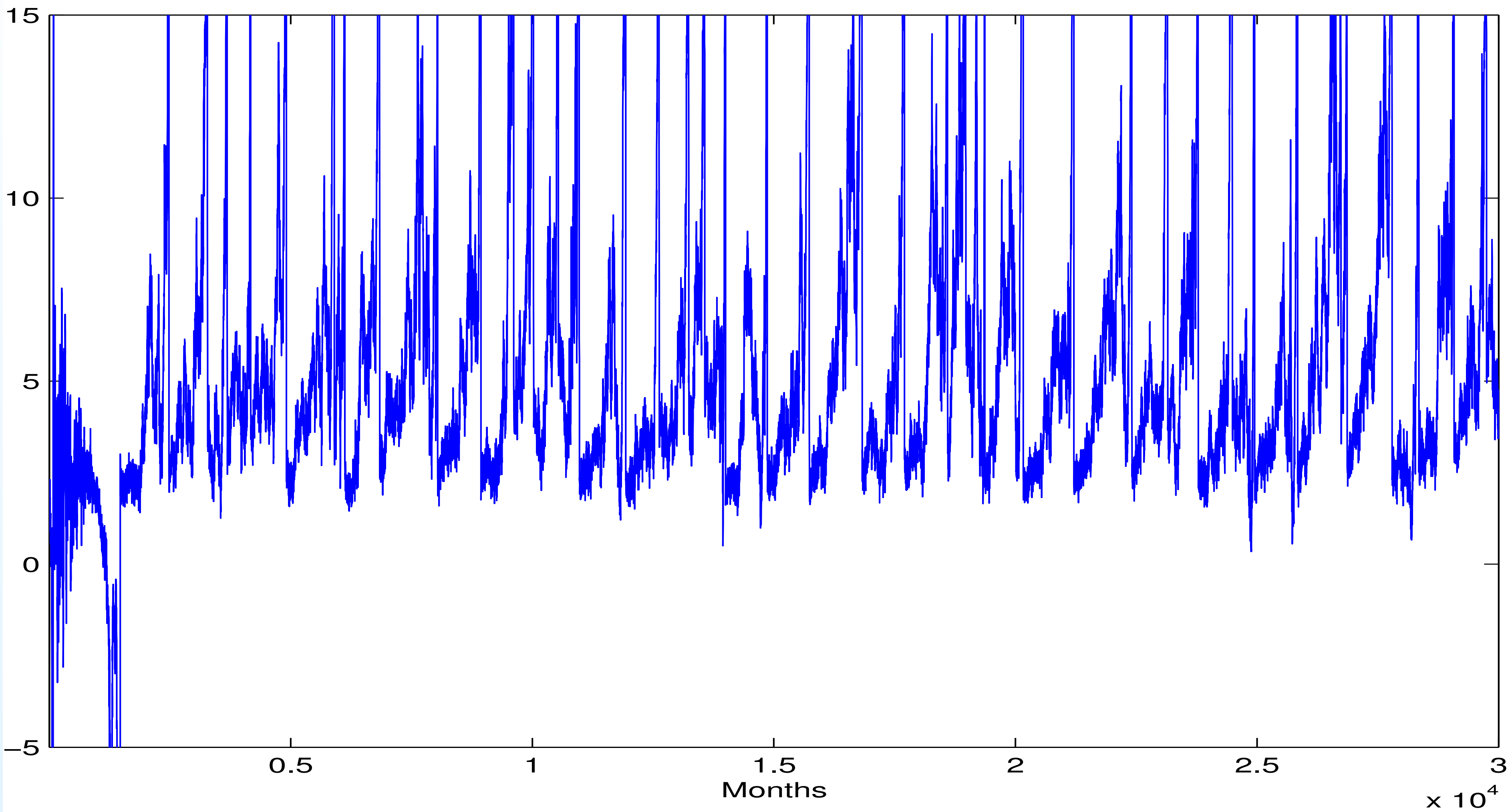
Dynamic forecasts with 68% and 90% error bands, using initial estimated condition at 77:01

Two peaks and an enduring decline (continued)



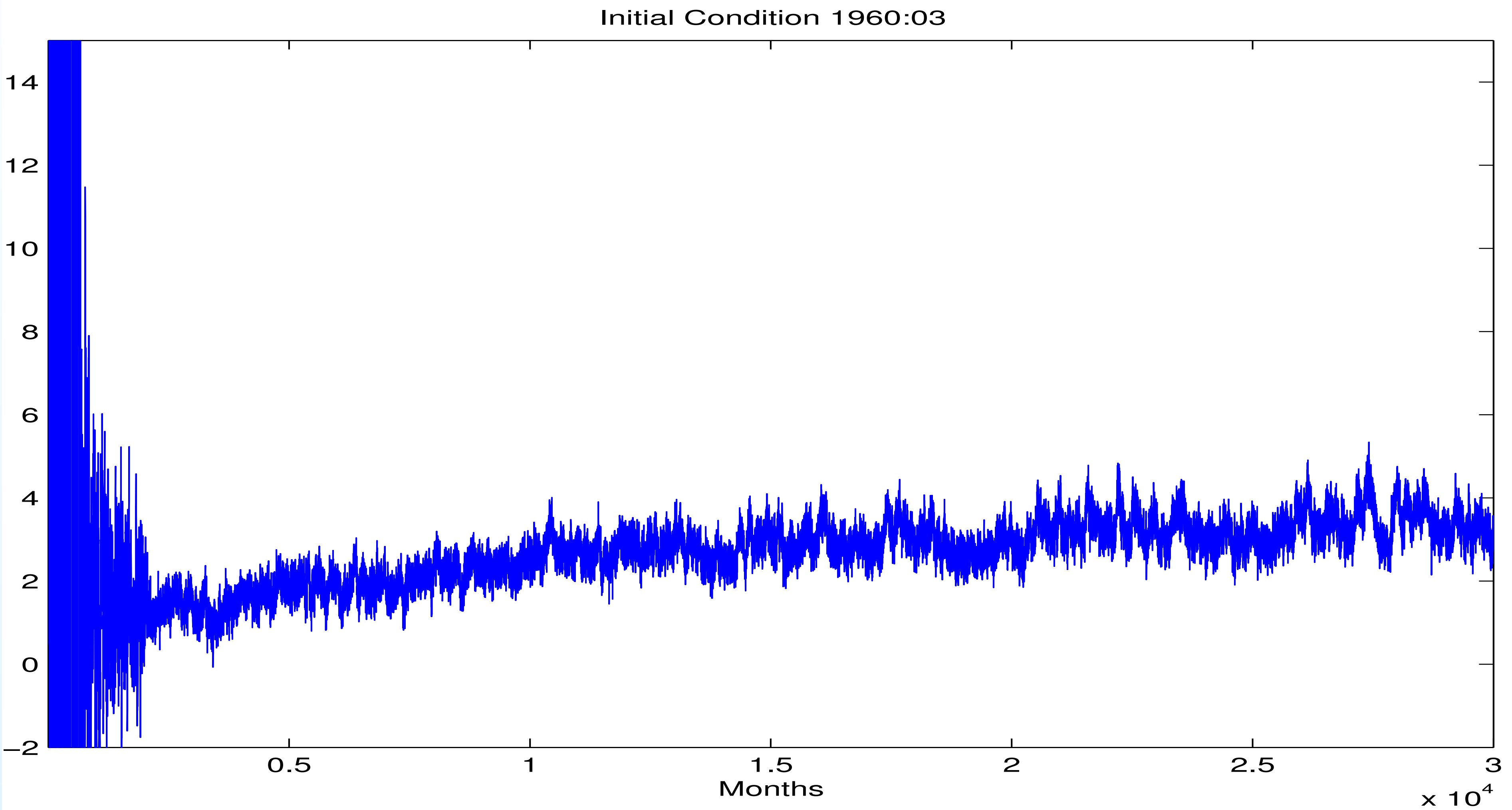
Dynamic forecasts with 68% and 90% error bands, using initial estimated condition at 80:01

SCE and escapes to the Ramsey equilibrium



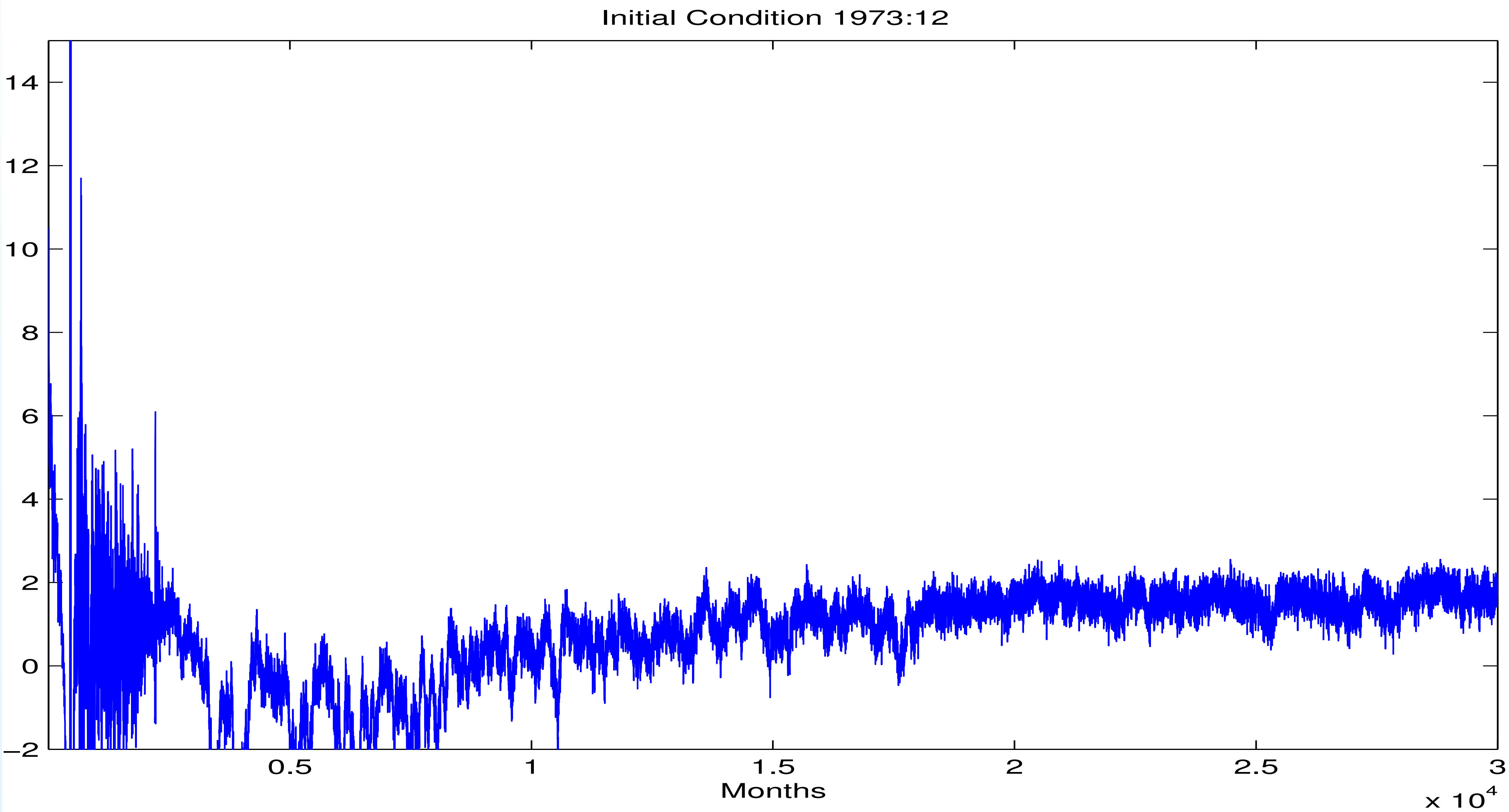
Frequent escapes with $\theta_0 = -1.0$, using the end-of-sample estimate as the initial condition

SCE and escapes to the Ramsey equilibrium (continued)



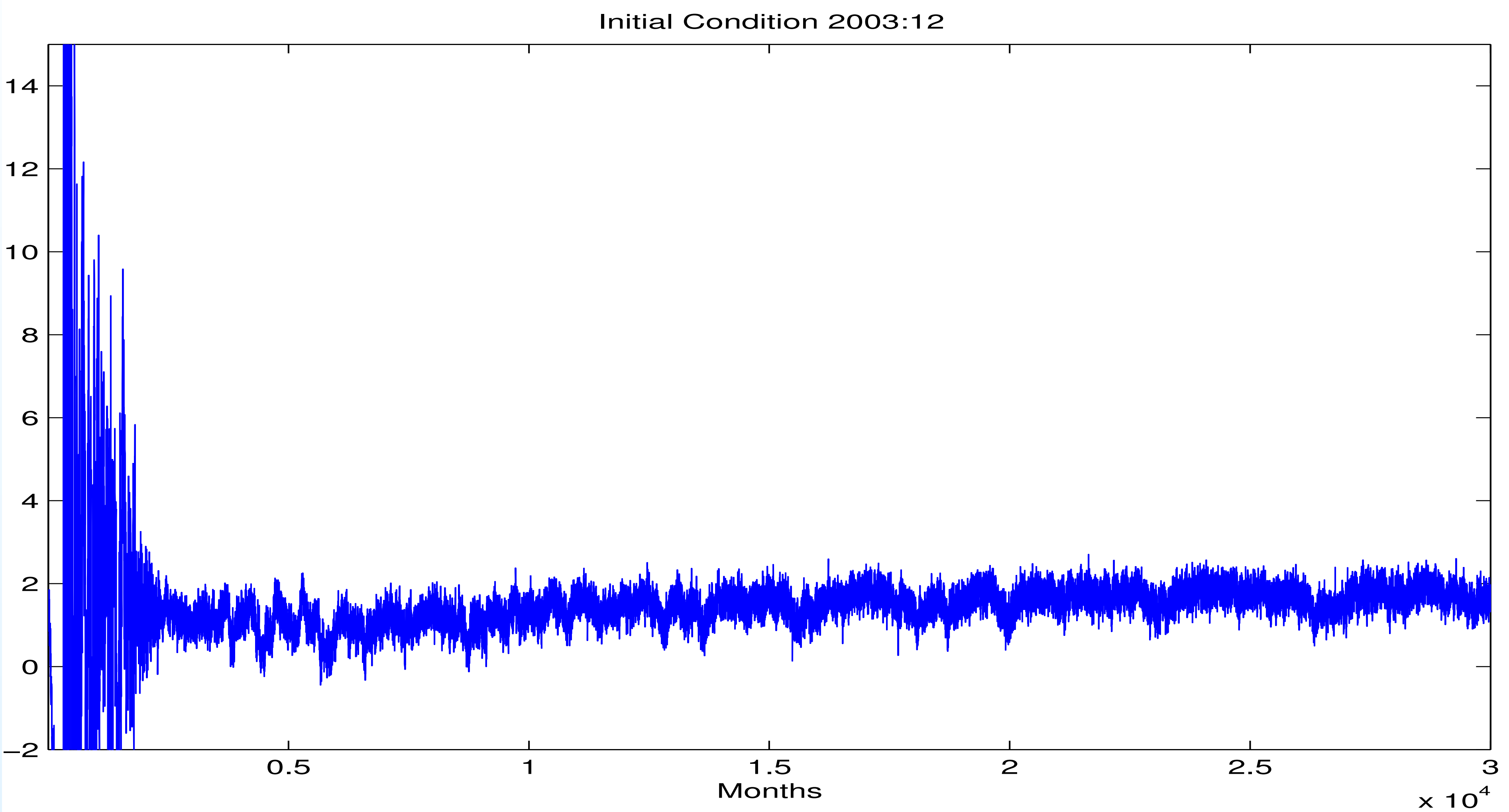
Government's inflation choice in one long Monte Carlo simulation, using estimated initial condition at 60:03

SCE and escapes to the Ramsey equilibrium (continued)



Government's inflation choice in one long Monte Carlo simulation, using estimated initial condition at 73:12

SCE and escapes to the Ramsey equilibrium (continued)



Government's inflation choice in one long Monte Carlo simulation, using estimated initial condition at 03:12

Model II: SZW hyperinflation model

- A hidden Markov, adaptive expectations model.
- The likelihood function.
- Escapes, mean dynamics, rational expectations, and SCE.
- Quantitative implications of our estimates for rational expectations and SCE versions.

Data

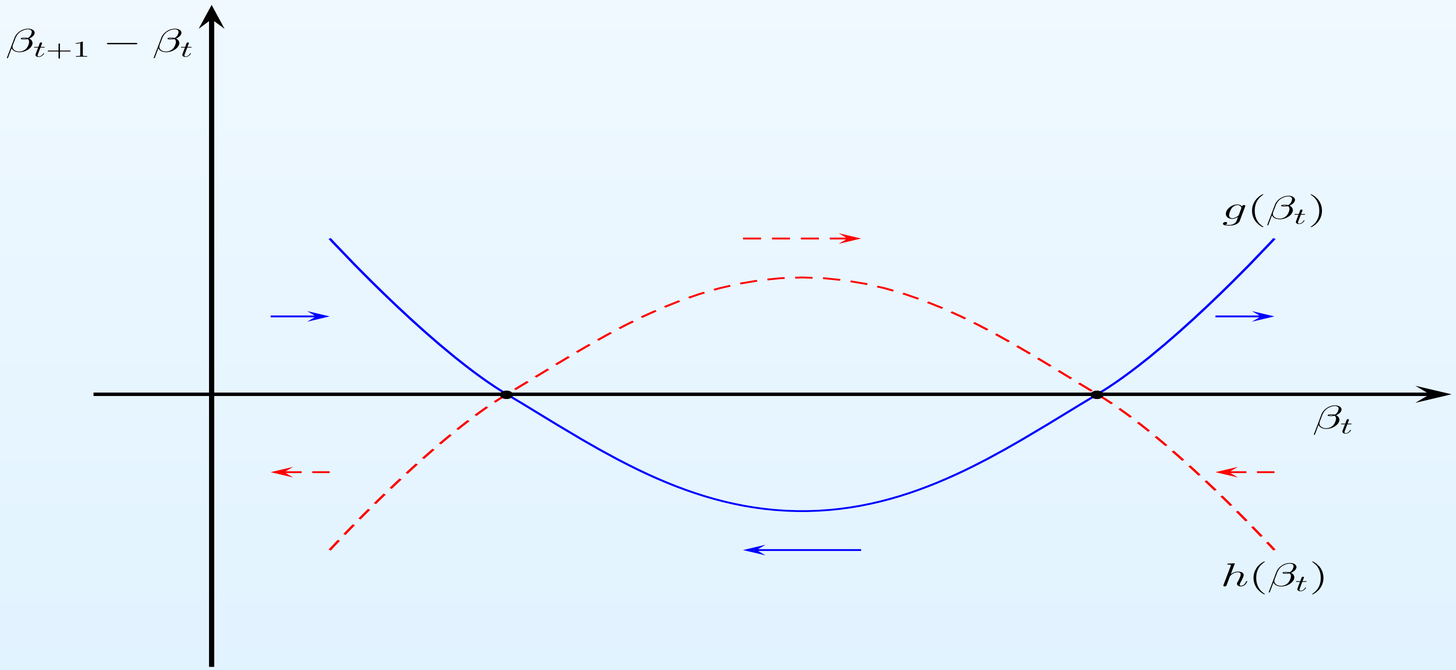
- Our estimates condition on data for monthly inflation only.
- We ignore data on deficits, but estimate the seigniorage component of the government's budget constraint.
- We form likelihood function for histories of inflation.
- The likelihood function is numerically challenging.

Literature

- Sargent and Wallace (1987) – rational expectations dynamics are perverse.
- Marcet and Sargent (1989) – least squares learning dynamics are more sensible.
- Marcet and Nicollini (2003) add an escape clause to the Marcet-Sargent model and calibrate it to explain recurrent hyperinflations.
- Escape dynamics of Sargent (1999), Cho, Williams, and Sargent (2002), and Williams (2005).
- Sargent, Williams, and Zha (2005) model of escape dynamics and U.S. inflation.

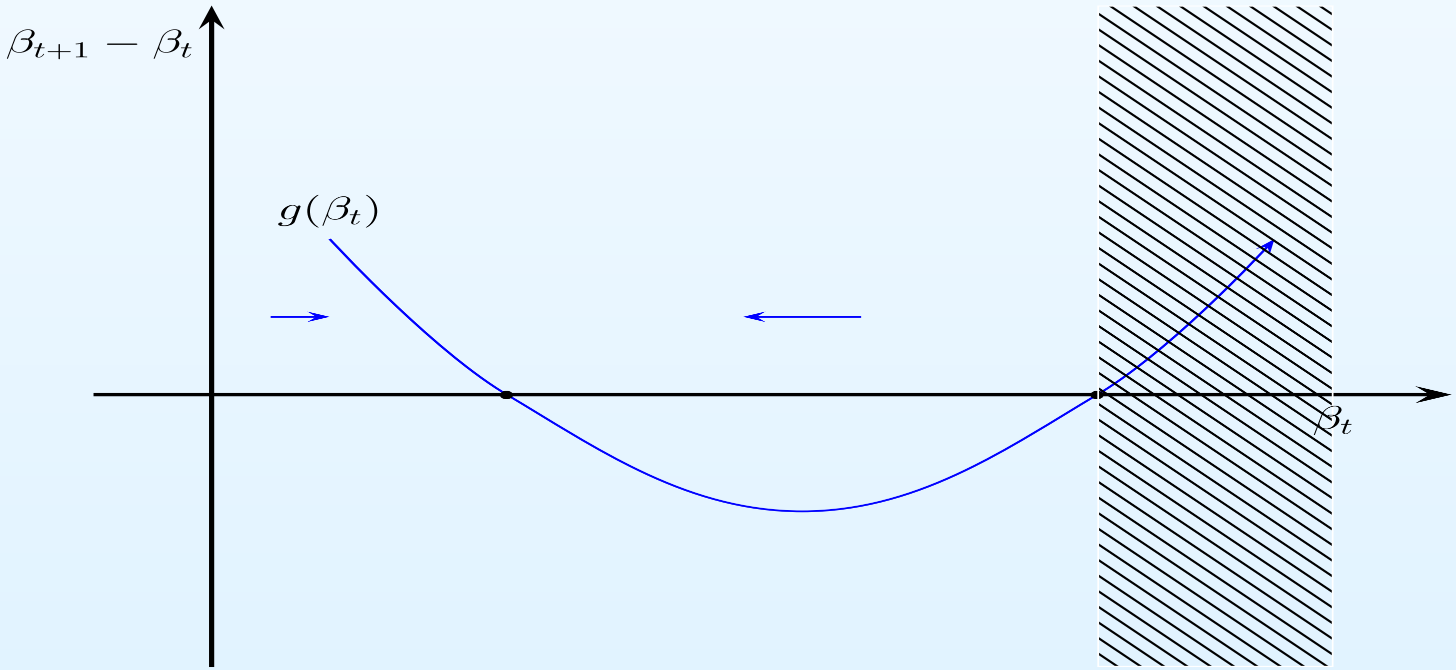
Adaptive and REE dynamics

Mean adaptive dynamics and REE dynamics.



Adaptive dynamics and escapes (continued)

Adaptive dynamics and the ‘escape event’.



The model, part 1

$$\frac{M_t}{P_t} = \frac{1}{\gamma} - \frac{\lambda}{\gamma}\frac{P^e_{t+1}}{P_t}, \tag{38}$$

$$M_t = \theta M_{t-1} + d_t(s_t)P_t, \tag{39}$$

$$d_t(s_t) = \bar{d}(s_t) + \eta_{d\,t}(s_t), \tag{40}$$

$$\text{Pr}(s_{t+1} = i | s_t = j) = q_{ij}, \; i, j = 1, ..., h, \tag{41}$$

$$\pi_{t+1} = \frac{P_{t+1}}{P_t}, \; \pi^b_{t+1} = \beta_t$$

The above model is a special case of (1) where $y_t = \pi_t$, $z_t = M_t$, and ϵ_t is a shock to $d_t(s_t)$.

$$\beta_t = \beta_{t-1} + g(\pi_{t-1} - \beta_{t-1}), \tag{42}$$

The learning rule (42) is a special case of (8) and (9), where $\alpha_{t|t-1} = \beta_t$, $x_{t-1} = 1$, and $P_{t|t-1} = 1$. Clearly, eqn:belief-cg is also a special case of (4).

No-breakdown conditions

$$1 - \lambda\beta_{t-1} > 0, \tag{43}$$

$$1 - \lambda\beta_t - d_t(s_t) > \delta\theta(1 - \lambda\beta_{t-1}), \tag{44}$$

The model, part 2

Expectations are reset after threatened breakdowns:

$$\pi_t = \pi_t^*(s_t) \equiv \pi_1^*(s_t) + \eta_{\pi\, t}(s_t), \tag{45}$$

$$p_\pi\left(\eta_{\pi\, t}(k)\right) = \begin{cases} \dfrac{\exp\left(-\frac{[\log\left(\pi_1^*(k)+\eta_{\pi\, t}(k)\right)-\log \pi_1^*(k)]^2}{2\sigma_\pi^2}\right)}{\sqrt{2\pi}\sigma_\pi\left(\pi_1^*(k)+\eta_{\pi\, t}(k)\right)\Phi((-\log(\delta)-\log(\pi_1^*(k))/\sigma_\pi)} & \text{if } -\pi_1^*(k) < \eta_{\pi\, t}(k) < 1/\delta - \pi_1^*(k) \\ 0 & \text{otherwise} \end{cases}, \tag{46}$$

$$p_d\left(\eta_{d\, t}(k)\right) = \begin{cases} \dfrac{\exp\left(-\frac{[\log\left(\bar d(k)+\eta_{d\, t}(k)\right)-\log \bar d(k)]^2}{2\sigma_d^2(k)}\right)}{\sqrt{2\pi}\sigma_d(k)\left(\bar d(k)+\eta_{d\, t}(k)\right)} & \text{if } \eta_{d\, t}(k) > -\bar d(k) \\ 0 & \text{if } \eta_{d\, t}(k) \le -\bar d(k) \end{cases}, \tag{47}$$

The escape event and its probability

Let

$$\varpi_t(k) = 1 - \lambda\beta_t - \delta\theta(1 - \lambda\beta_{t-1}) - \bar{d}(k),$$

for $k = 1, \ldots, h$. The probability of *escape* at time t is

$$\mathcal{I}(\beta_{t-1} < 1/\lambda) \sum_{k=1}^{h} \Big[\Pr(s_t = k | \Pi_{t-1}, \hat{\phi}) \int_{-\bar{d}(k)}^{\varpi_t(k)} \mathcal{I}(\pi_2^*(k) < \beta_t) p_d(\eta_{d\,t}(k)) d\,\eta_{d\,t}(k) \Big].$$

- If perceived inflation β_t is above $\pi_2^*(s_t)$, actual inflation is on average higher than β_t, perceived inflation tends to escalate, and hyperinflation is likely to occur – this is the escape event.
- If perceived inflation β_t is above $\varpi_t(s_t)$, however, equilibrium is ϵ away from breaking down and mechanical “reforms” on expectations take place.

The breakdown event and its probability

The probability of *breakdown* at time t is

$$\mathcal{I}(\beta_{t-1} \geq 1/\lambda) + \mathcal{I}(\beta_{t-1} < 1/\lambda) \sum_{k=1}^{h} \Big[\text{Pr}(s_t = k | \Pi_{t-1}, \hat{\phi}) \\ \int_{\varpi_t(k)}^{\infty} p_d(\eta_{d\,t}(k)) d\,\eta_{d\,t}(k) \Big].$$

- If $\beta_{t-1} \geq 1/\lambda$, equilibrium at $t-1$ was broken down and real balances was negative. In this situation, the probability of no equilibrium at t becomes one.
- If real balances at $t-1$ is positive and perceived inflation β_t is above $\varpi_t(s_t)$, we are entering the territory threatening the existence of a positive price level to support the equilibrium.

The likelihood function

The conditional likelihood is

$$
\begin{aligned}
p(\pi_t|\Pi_{t-1}, S_T, \phi) &= p(\pi_t|\Pi_{t-1}, s_t, \phi) \\
&= C_{1\,t}\, \frac{|\xi_\pi|\,\text{exp}\left\{-\frac{\xi_\pi^2}{2}\left(\text{log}\pi_t - \text{log}\pi_1^*(s_t)\right)^2\right\}}{\sqrt{2\pi}\;\Phi\left(|\xi_\pi|(-\text{log}(\delta) - \text{log}(\pi_1^*(s_t))\right)\;\pi_t} \\
&\quad + C_{2\,t}\left[\frac{\theta|\xi_d(s_t)|(1-\lambda\beta_{t-1})}{\sqrt{2\pi}\;[(1-\lambda\beta_t)\pi_t - \theta(1-\lambda\beta_{t-1})]\,\pi_t}\right. \\
&\quad \left.\text{exp}\left\{-\frac{\xi_d^2(s_t)}{2}\Big[\text{log}[(1-\lambda\beta_t)\pi_t - \theta(1-\lambda\beta_{t-1})] - \text{log}\pi_t - \text{log}d(s_t)\Big]^2\right\}\right],
\end{aligned}
\tag{48}
$$

The likelihood function (continued)

where

$$C_{1\,t} = \mathcal{I}(\beta_{t-1} \geq 1/\lambda) + \mathcal{I}(\beta_{t-1} < 1/\lambda)$$
$$\left(1 - \Phi\left[|\xi_d(s_t)|\left(\text{log}\left(\text{max}[(1-\lambda\beta_t) - \delta\theta(1-\lambda\beta_{t-1}), 0]\right) - \text{log} d(s_t)\right)\right]\right),$$
$$C_{2\,t} = \mathcal{I}(\beta_{t-1} < 1/\lambda)\,\mathcal{I}\left(\frac{\theta\,(1-\lambda\,\beta_{t-1})}{\text{max}\,(1-\lambda\,\beta_t,\ \delta\theta(1-\lambda\,\beta_{t-1}))} < \pi_t < \frac{1}{\delta}\right).$$

Estimation

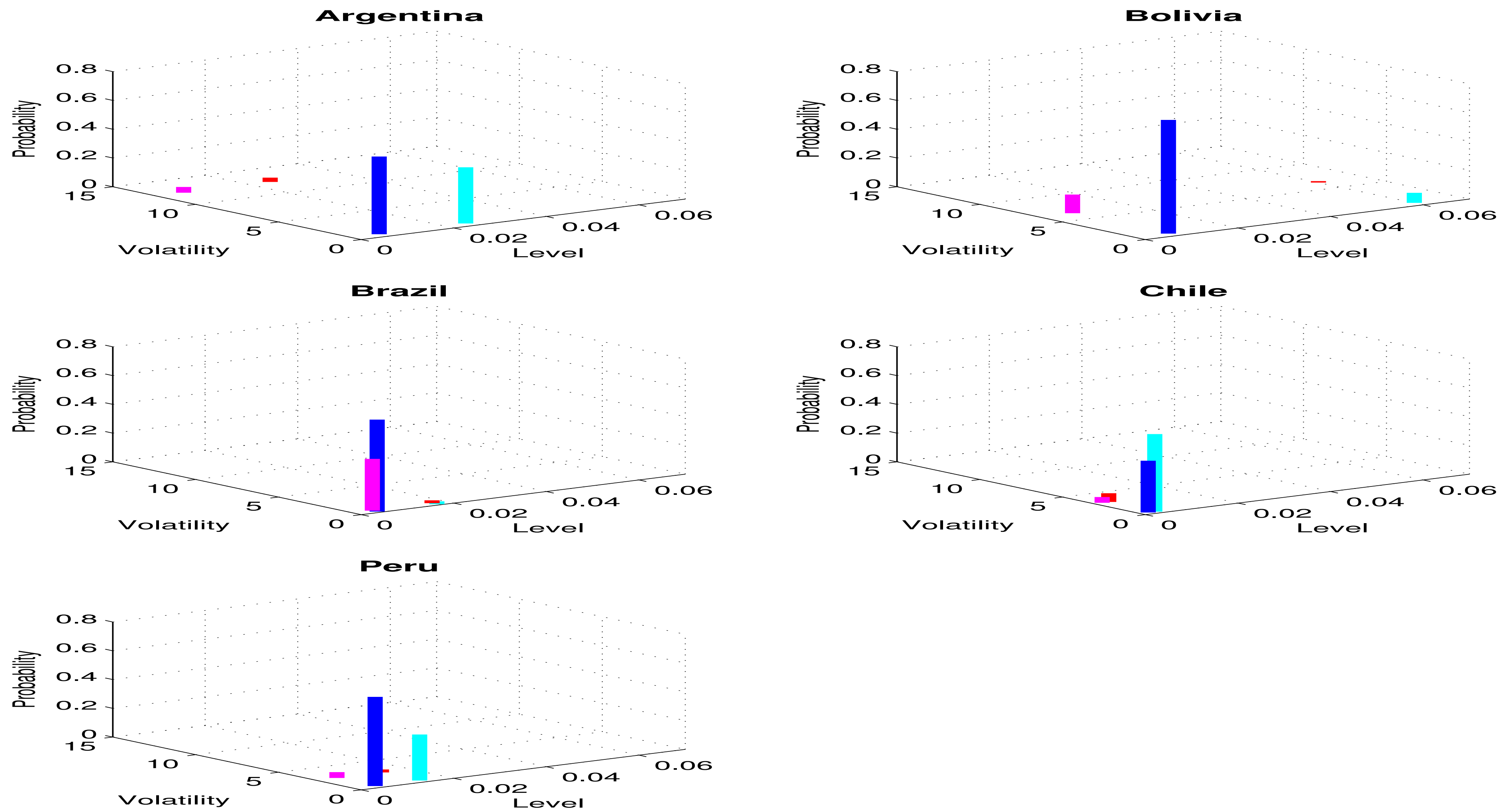
- The overall likelihood can be computed recursively. It is quite complicated and has local peaks.
- Estimation: (1) use the block-wise BFGS algorithm following the idea of Gibbs sampling and EM algorithm; (2) iterate between this BFGS algorithm and the IMSL constrainted optimization routine.
- Estimation: (1) start with a grid of 300 starting points and perturb around each local peak point in both small and big steps to generate additional 200 new starting points; (2) utilize the parallel and grid computing technology on the Linux OS; (3) it takes 5 days to get the MLE for one country.

Empirical results: log likelihood

Table 1: Log likelihood adjusted by the Schwarz criterion

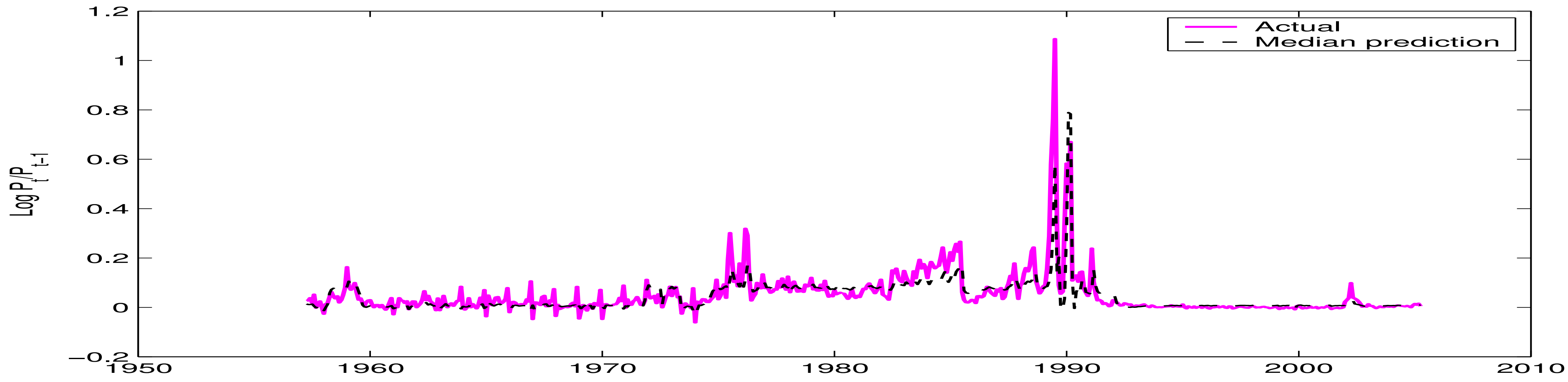
	Hyperinflation Model	2-state AR(2) (df=-1)	2×2-state AR(2) (df=2)
Argentina	1232.5	1095.2	1490.1
Bolivia	1505.6	1483.7	1539.9
Brazil	750.34	782.97	838.02
Chile	1697.3	1605.4	1714.6
Peru	1651.0	1517.1	1652.3

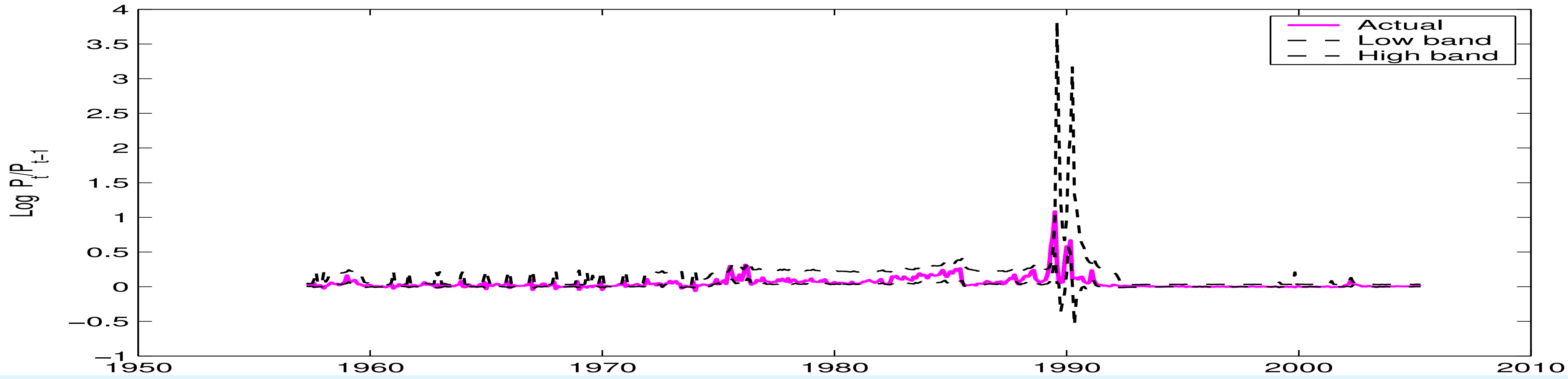
Empirical results: probabilities of regimes



Ergodic probability given the estimated seigniorage level (x-axis) and the estimated standard deviation of shock (y-axis).

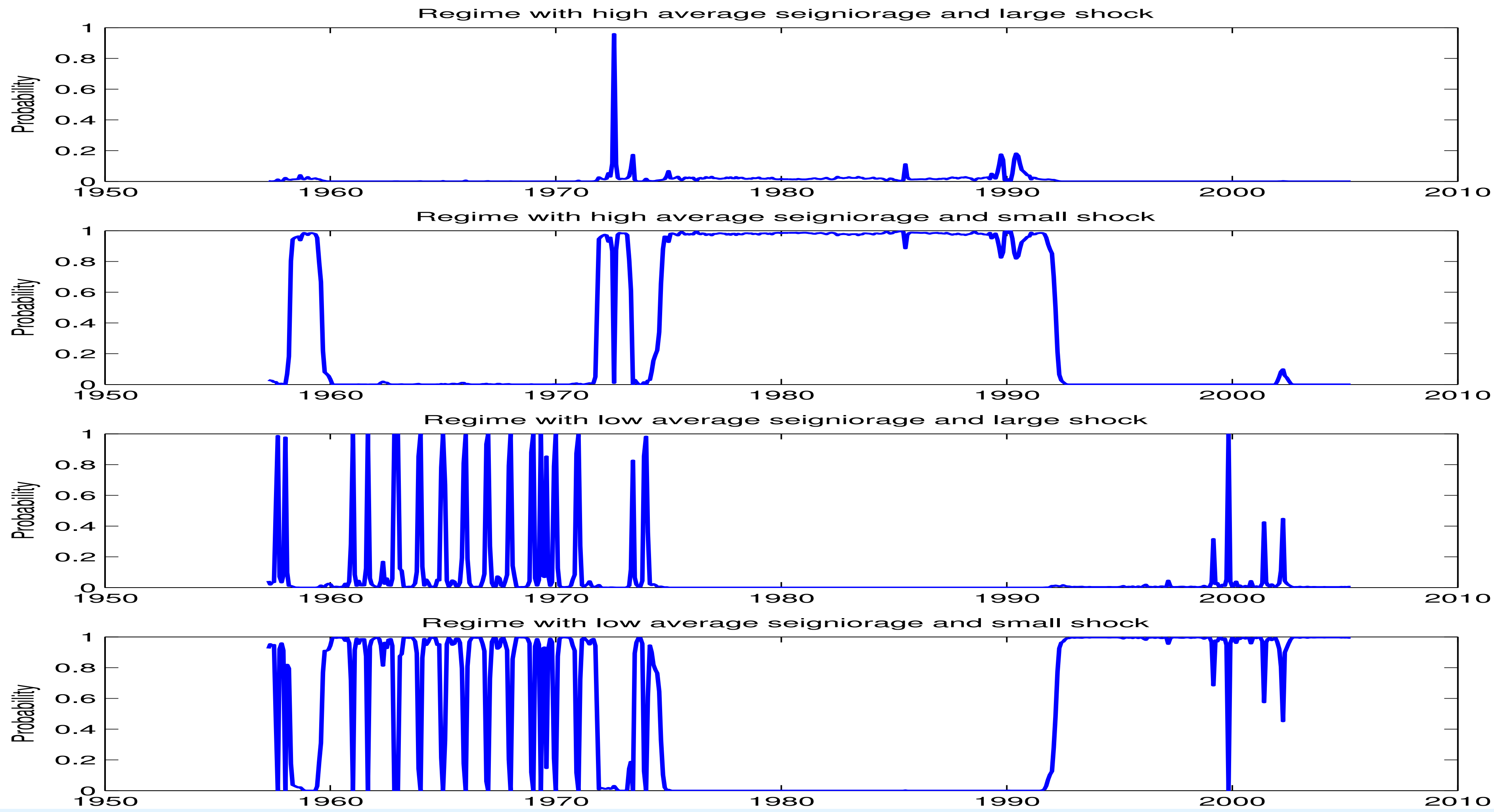
Empirical results: Argentina





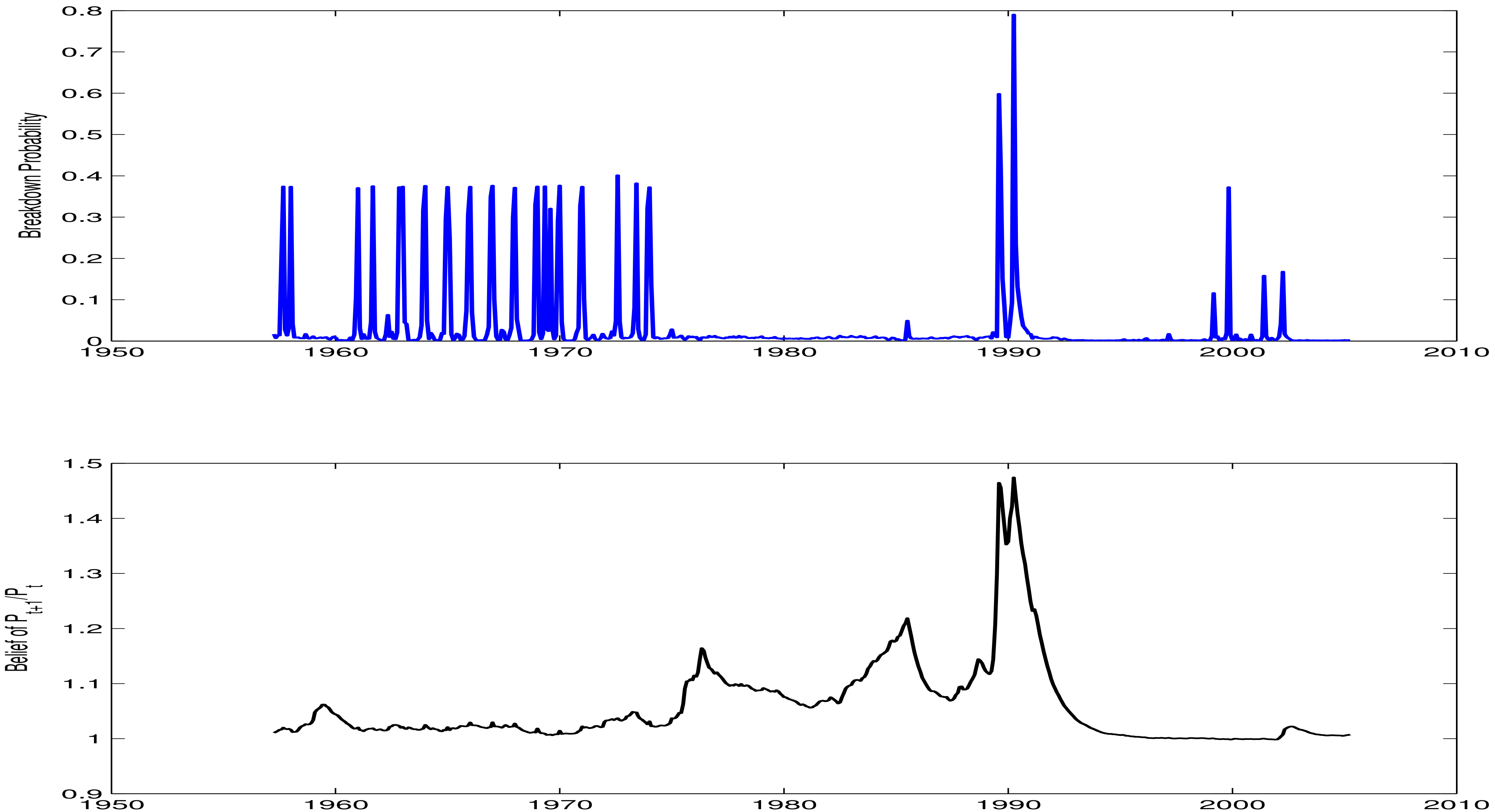
Argentinean inflation: actual versus one-step median forecast (top chart) and actual versus .90 probability bands of one-step prediction (bottom chart).

Empirical results: Argentina (continued)



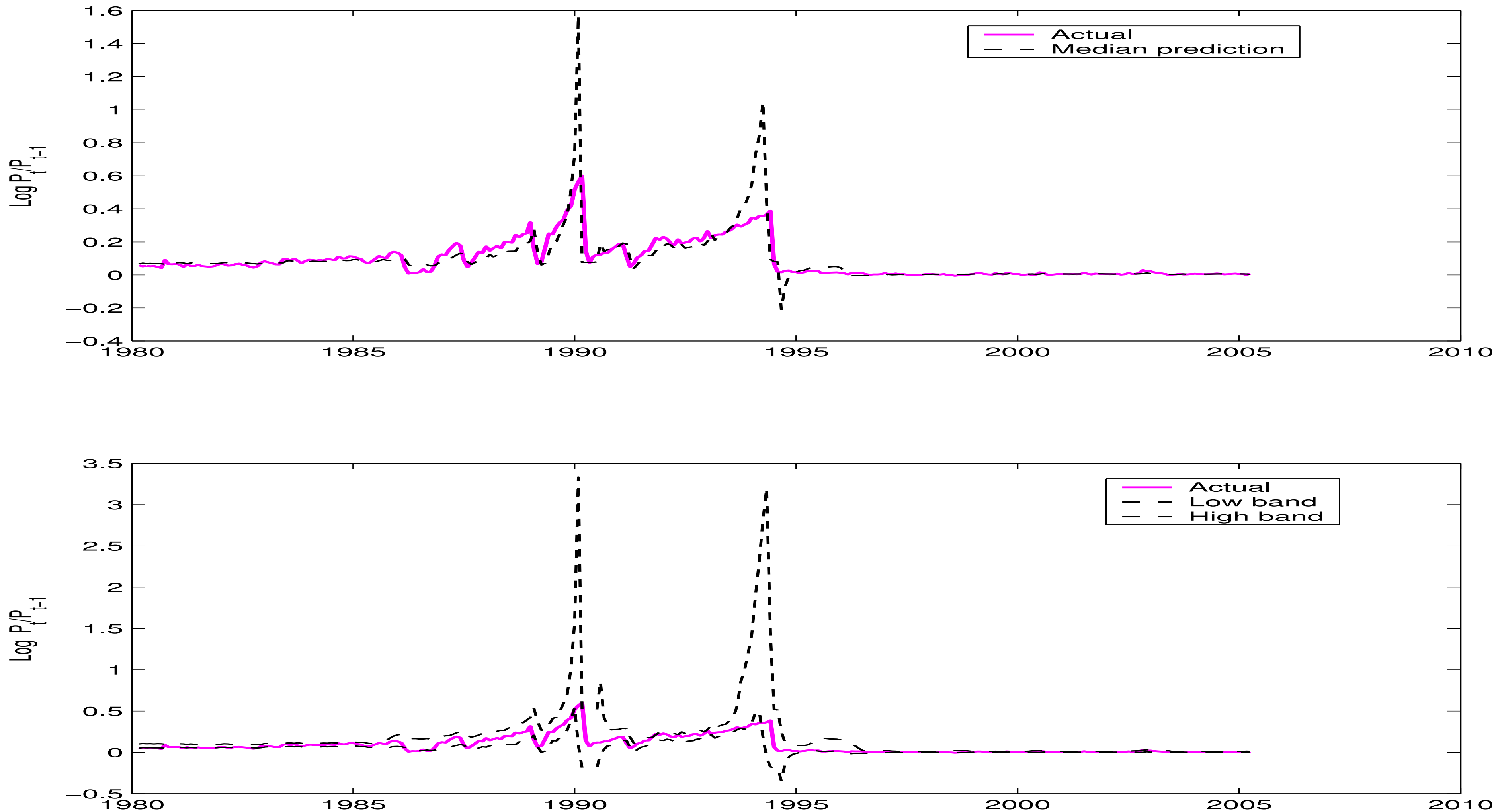
Argentina: probabilities of the four regimes conditional on the MLEs and the data.

Empirical results: Argentina (continued)



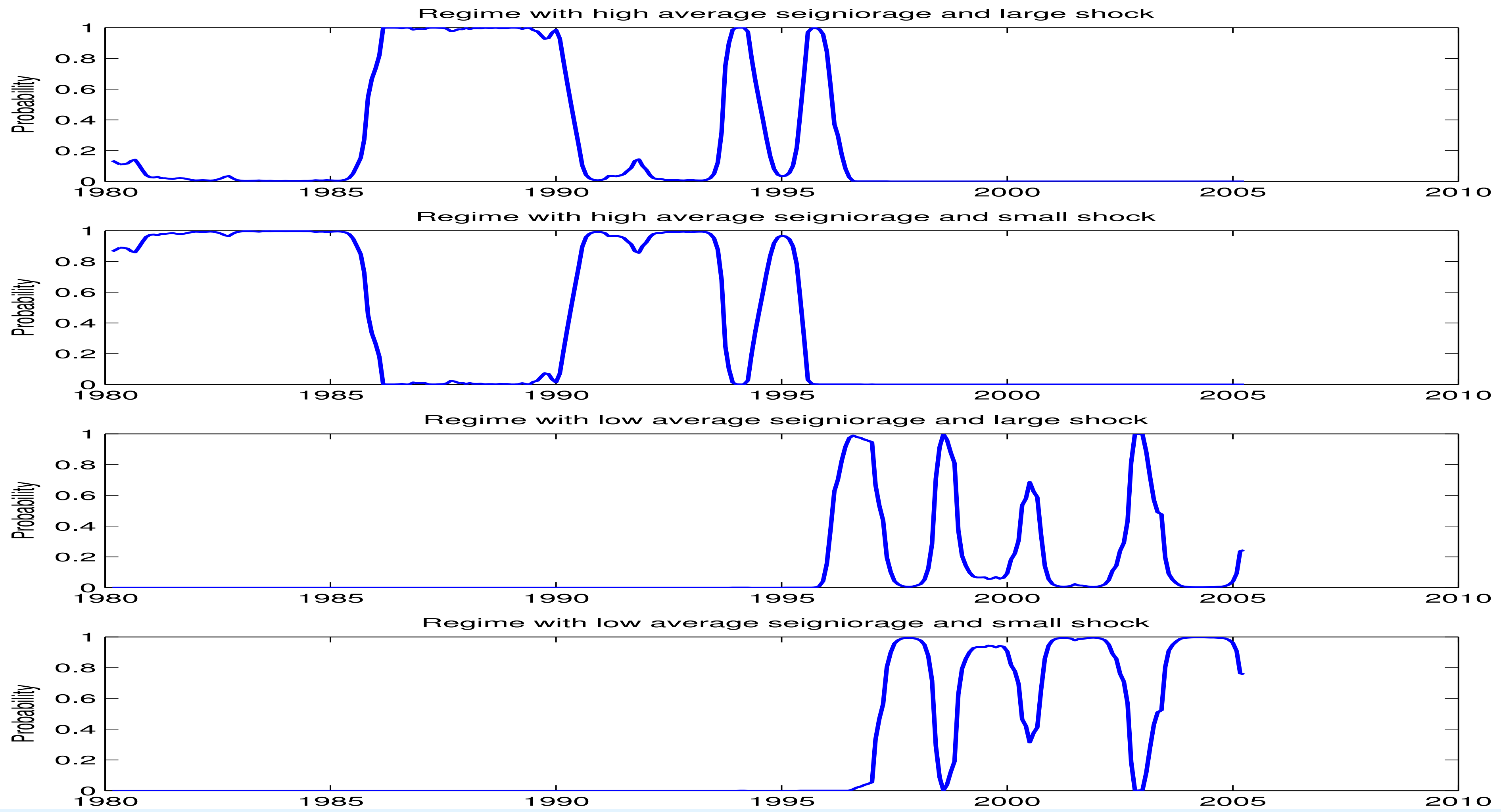
Argentina: probability of breakdown (top chart) and belief of next-period inflation β_t (bottom chart).

Empirical results: Brazil



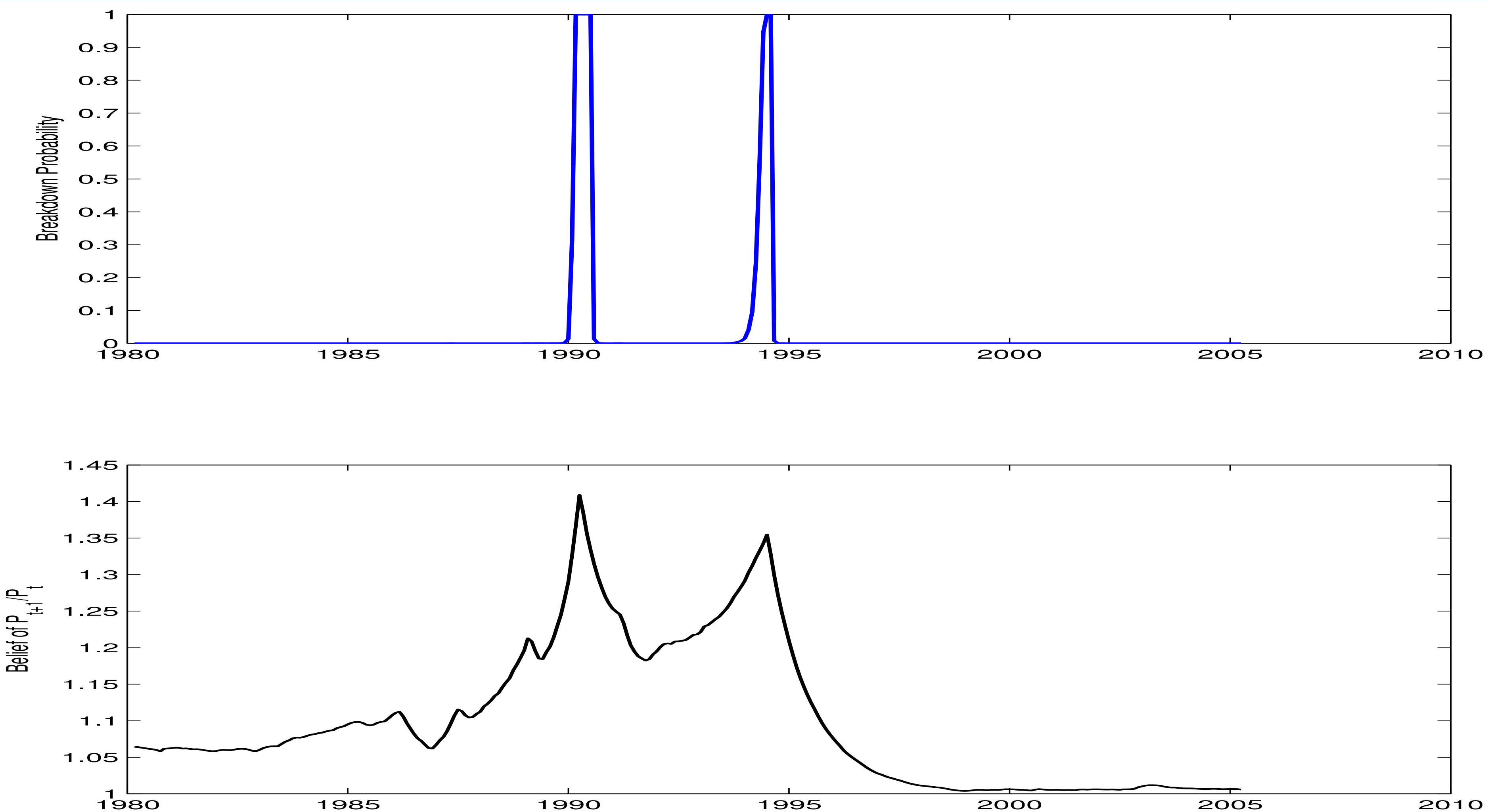
Brazilian inflation: actual versus one-step median forecast (top chart) and actual versus .90 probability bands of one-step prediction (bottom chart).

Empirical results: Brazil (continued)



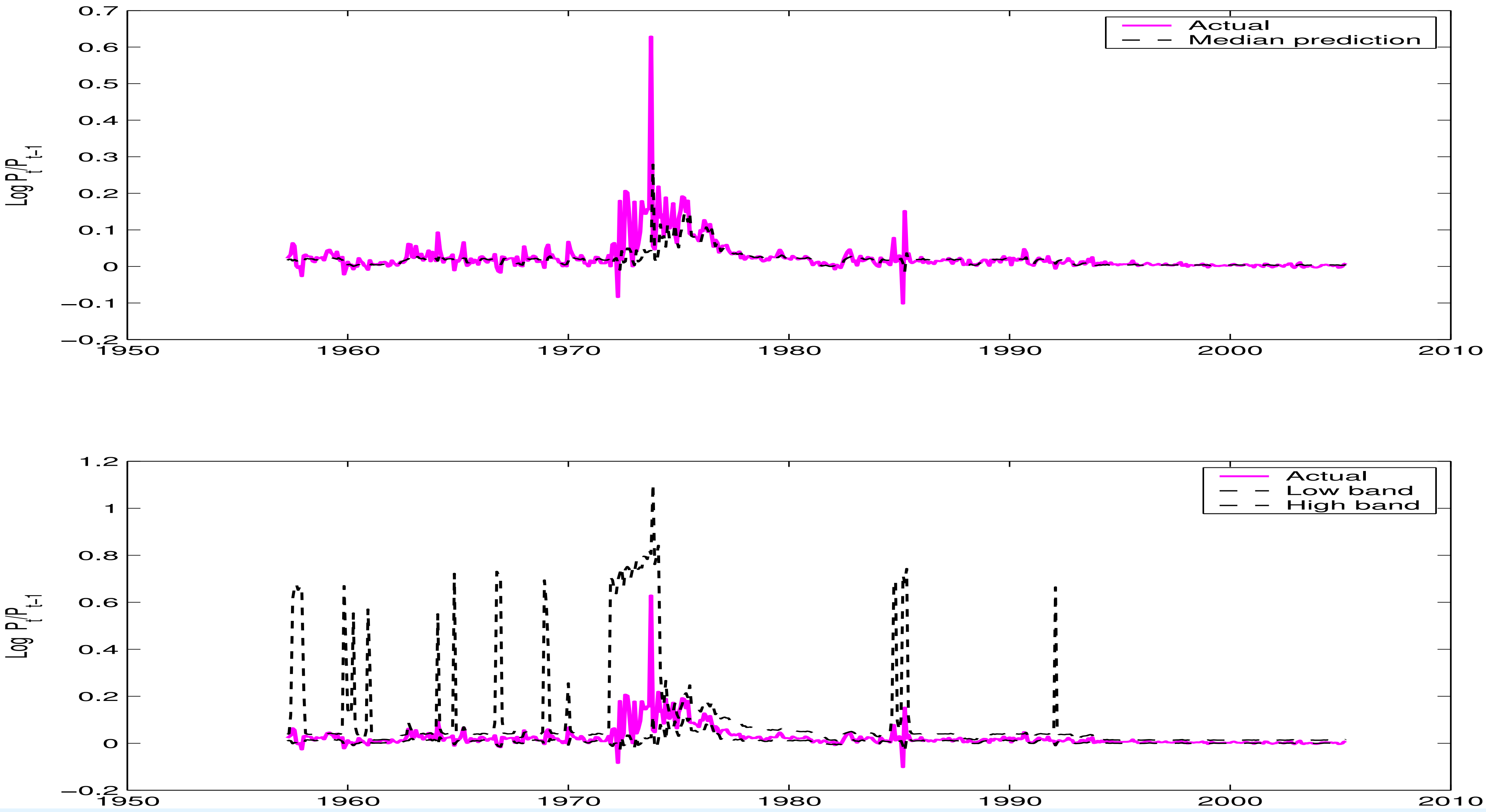
Brazil: probabilities of the four regimes conditional on the MLEs and the data.

Empirical results: Brazil (continued)



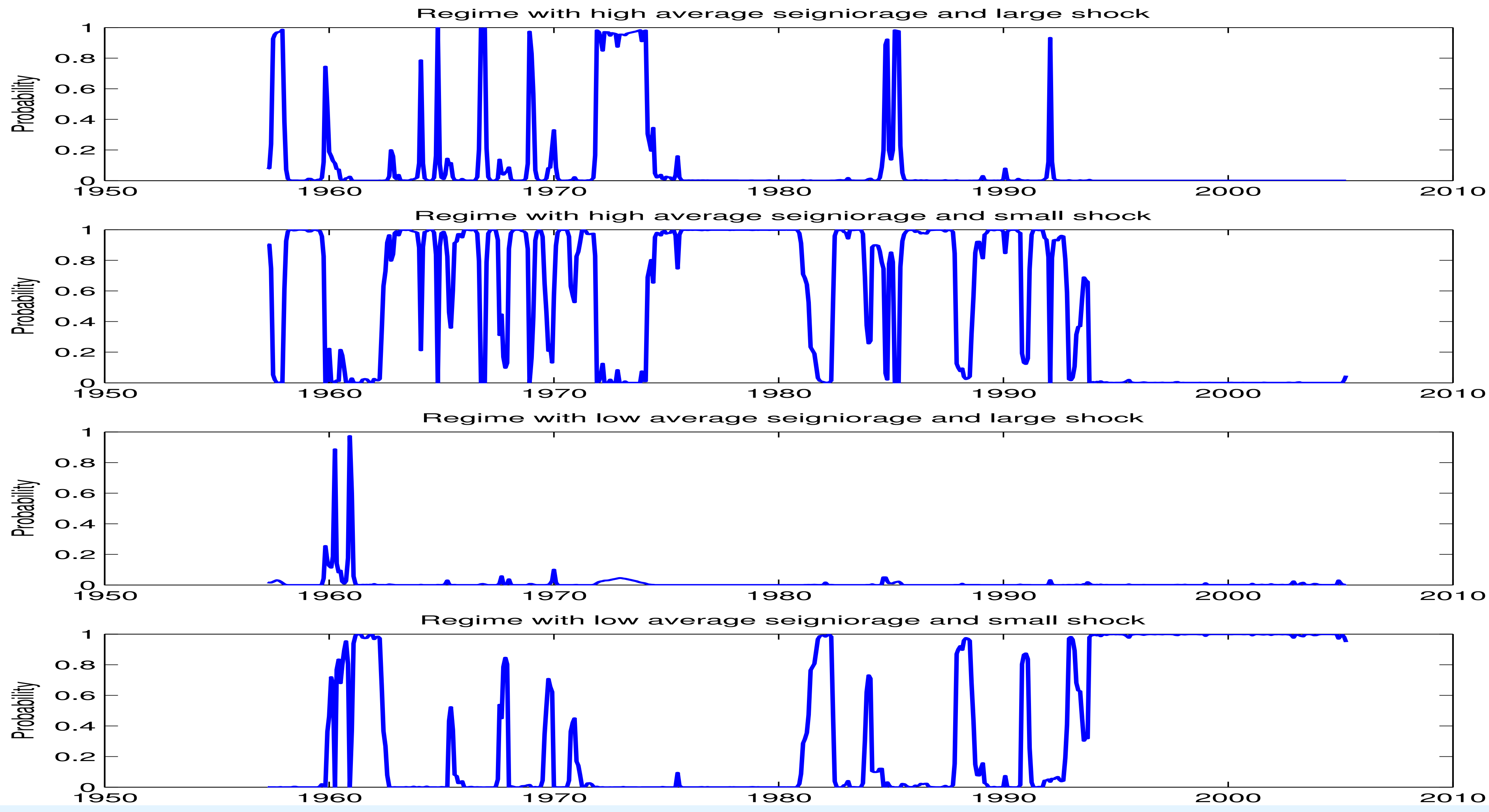
Brazil: probability of breakdown (top chart) and belief of next-period inflation β_t (bottom chart).

Empirical results: Chile



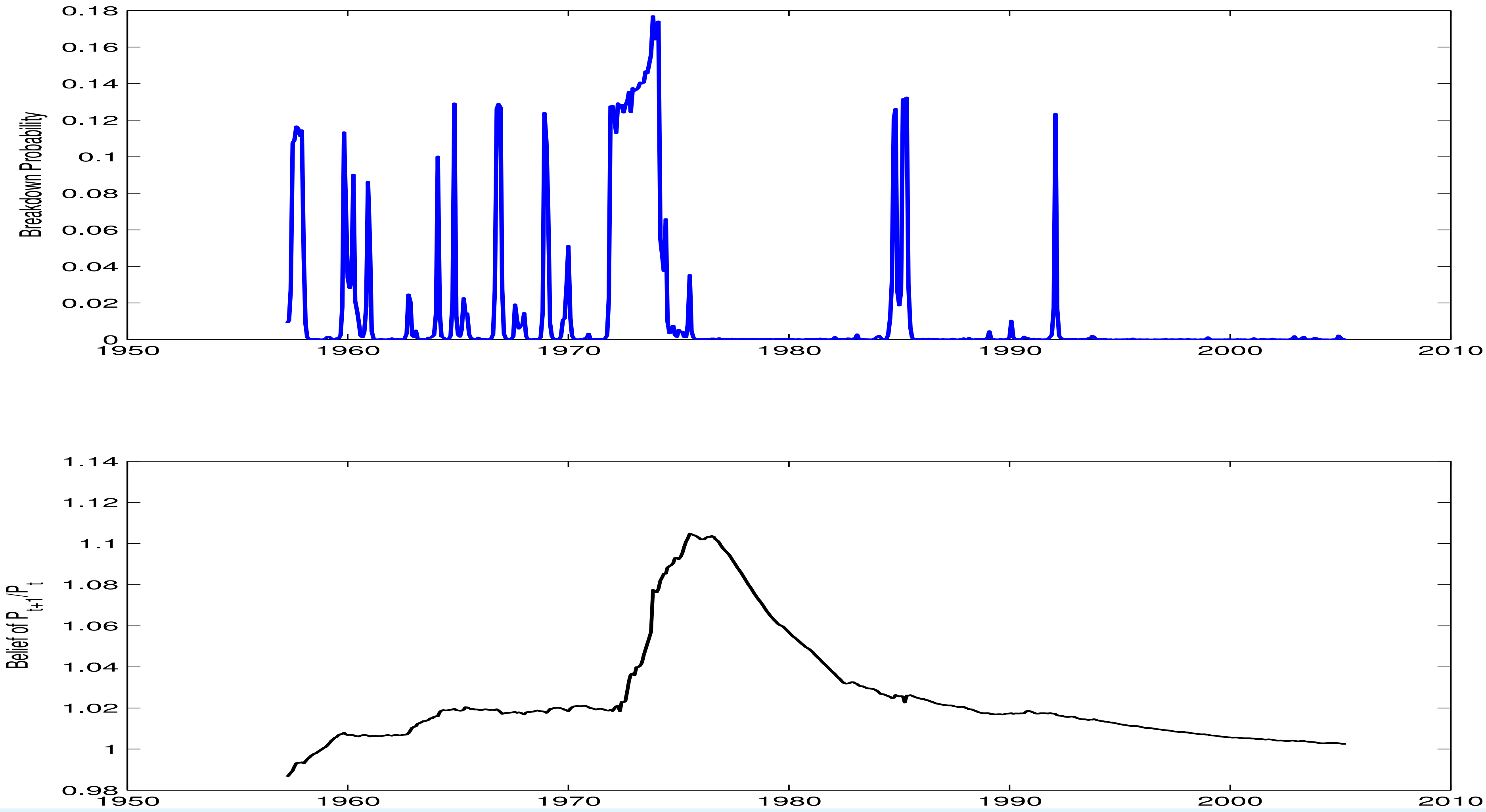
Chilean inflation: actual versus one-step median forecast (top chart) and actual versus .90 probability bands of one-step prediction (bottom chart).

Empirical results: Chile (continued)



Chile: probabilities of the four regimes conditional on the MLEs and the data.

Empirical results: Chile (continued)



Chile: probability of breakdown (top chart) and belief of next-period inflation β_t (bottom chart).

Other theories of expectations and equilibrium

- Rational expectations. Note that $\pi_t = \pi(s_t, s_{t-1}, d_t)$ and denote $\pi^e(s_t) \equiv E_t \pi_{t+1}$. It can be shown that π_t is a nonlinear function of $\pi^e(s_t)$ and $\pi^e(s_{t-1})$, denoted by $h\left(\pi^e(s_t), \pi^e(s_{t-1})\right)$. Specifically,

$$\pi(s_t, s_{t-1}, d_t) = h\left(\pi^e(s_t), \pi^e(s_{t-1})\right).$$

  Taking expectations of both sides conditional on information at $t-1$ with $s_{t-1} = i$ gives

$$\pi^e(i) = E_{t-1} h(\pi^e(s_t), \pi^e(i)).$$

  The numerical solution for $\pi^e(i)$ is to find a fixed point to the above nonlinear equation. The solution is not unique and there are multiple equilbria.

  The unconditional rational expectations of π_t is to take the average of $\pi^e(i)$ according to the ergodic probability of s_t.

Other theories of expectations and equilibrium

- Self-confirming equilibrium (SCE). According to (14) where $\alpha_{\text{SCE}} = \beta_{\text{SCE}}$, $x_t = 1$, $y_t = \pi_t$, we need to compute the fixed point β_{SCE} that solves

  $$E\pi_t(\beta_{\text{SCE}}) - \beta_{\text{SCE}} = 0,$$

  where π_t depends on β_{SCE}.

  For an SCE conditional on $s_t = i$, compute the fixed point $\beta_{\text{SCE}}(i)$ that solves

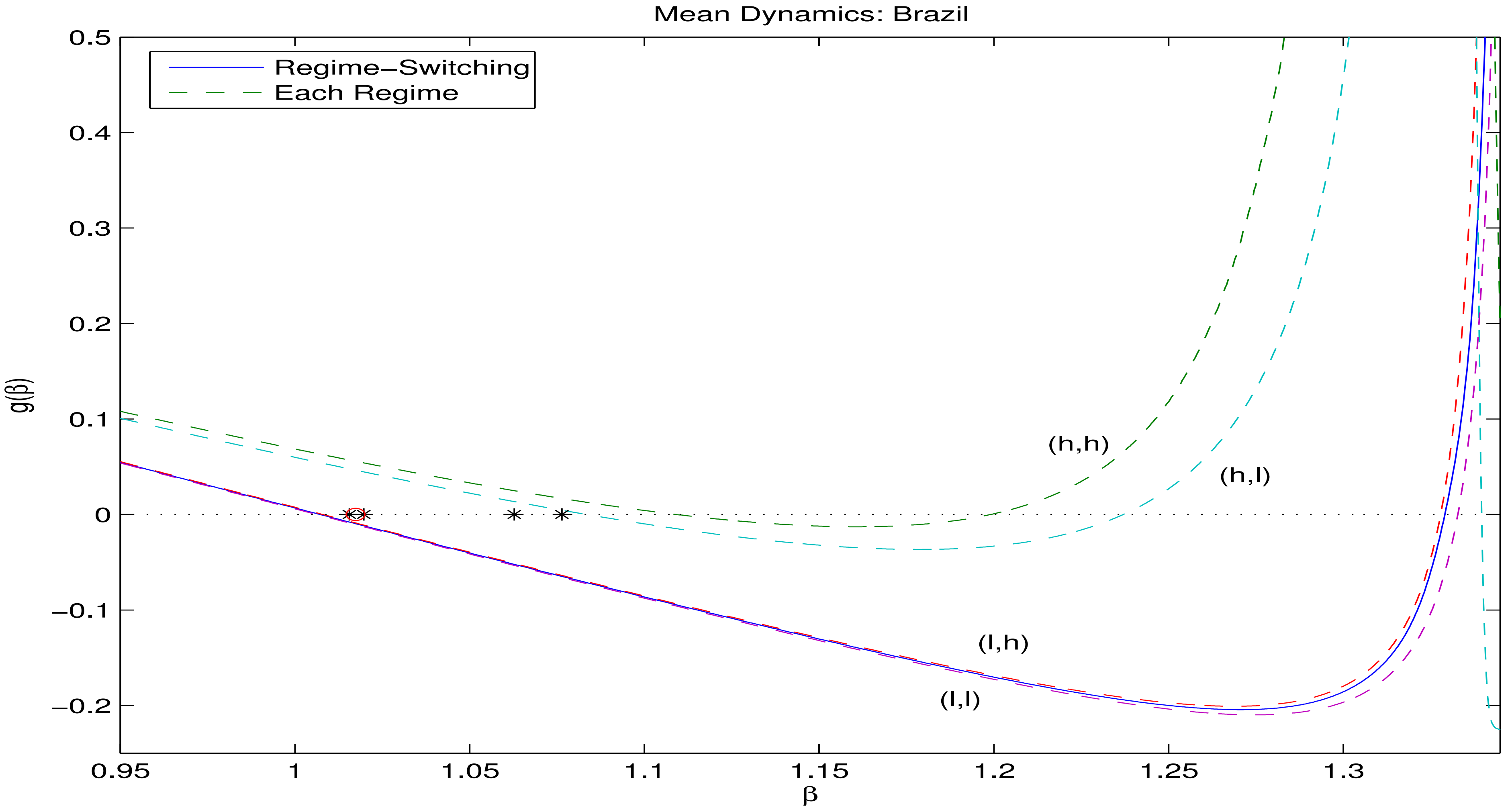
  $$E\left(\pi_t(\beta_{\text{SCE}}) \mid s_t = i\right) - \beta_{\text{SCE}}(i) = 0,$$

  where $E(\ |s_t = i)$ is the conditional expectation.
- Relationship between adaptive expectations and SCE: the mean dynamics and escape dynamics.
- We'll study these at our parameter estimates – a way of studying how big is our departure from rational expectations.

SCE and conditional SCE

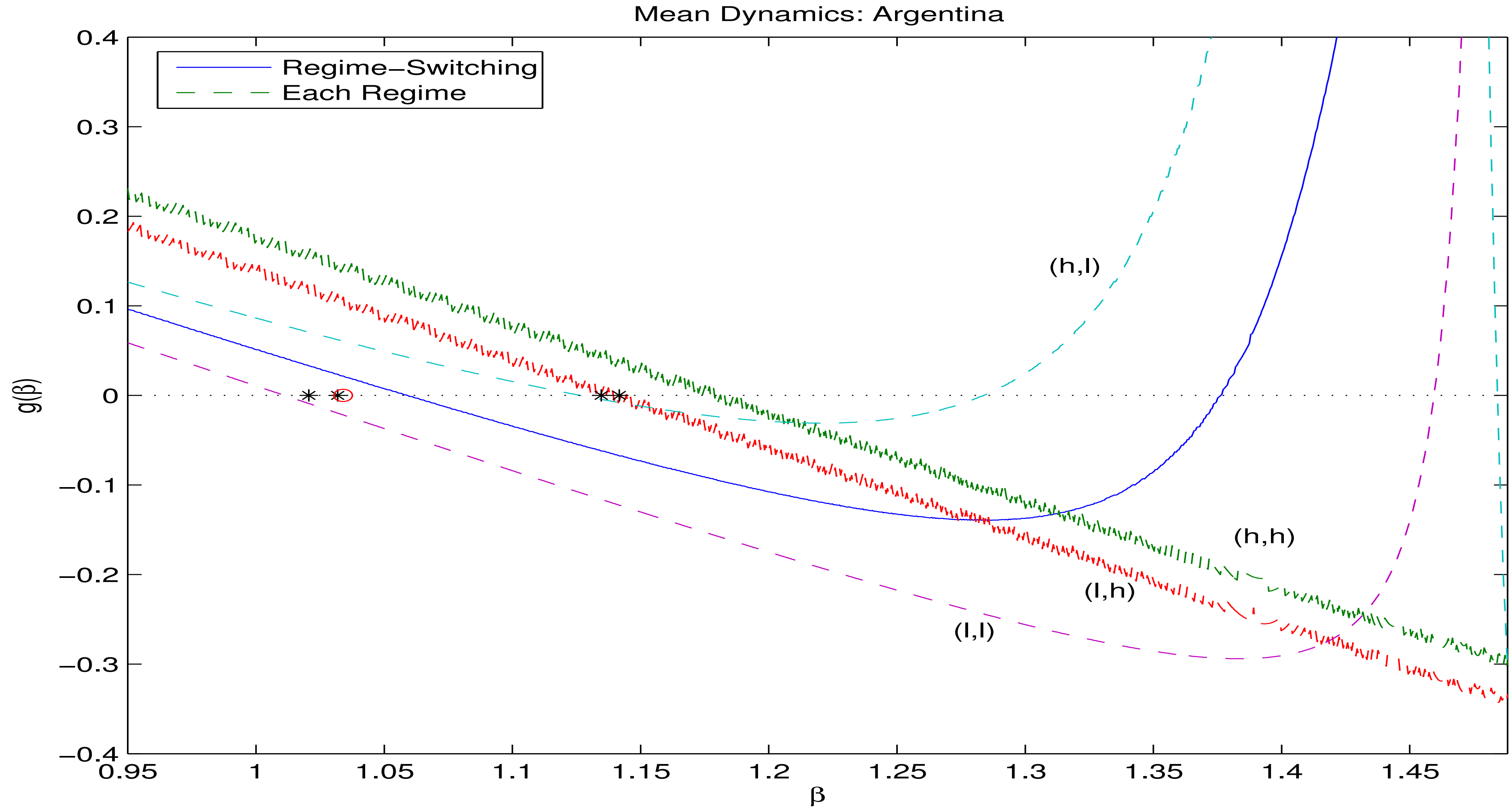
- Conditional SCE: pretend that regime is constant, compute self-confirming β.
- Self-confirming equilibrium (SCE): find mean dynamics for β, assuming regime switching, and compute fixed point.
- Compute REE forecasts of inflation in each state s_t.
- View conditional SCE's as approximation to REE forecasts of inflation conditional on state s_t.
- View SCE as an approximation to unconditional REE forecast of inflation.

Brazil SCE



Mean dynamics, conditional mean dynamics, and REE for Brazil.

Argentina SCE



Mean dynamics, conditional mean dynamics, and REE for Argentina.

Concluding remarks: the causes of inflation

- Our model attributes inflation to shocks, states that describe the average and the volatility of deficits, and beliefs.
- We use a model with adaptive expectations, endogenous escapes, and mechanical "reforms" that operate directly on expectations.
- Staying in the (ℓ, ℓ) regime is the key to arresting inflation.

Concluding remarks (continued)

- From our estimates, we can project what the deficit data should be.
- We have provided evidence that our adaptive expectations model often gives outcomes close to REE.
- As in Marcet-Nicolini, a small deviation from REE gives escape dynamics that can help explain recurrent big inflations in a way that gives a somewhat independent role to expectations.