The Conquest of U.S. Inflation: Learning and Robustness to Model Uncertainty

Timothy Cogley and Thomas Sargent

University of California, Davis and New York University and Hoover Institution

The Question

The Question

- After ample evidence had accumulated in the 1970s in favor the natural unemployment rate hypothesis, why did the Fed wait so long to bring inflation down?

The Question

- After ample evidence had accumulated in the 1970s in favor the natural unemployment rate hypothesis, why did the Fed wait so long to bring inflation down?
- ‘The picture’

Standard answers take for granted that the Fed always knew that the rational expectations version of the natural rate hypothesis was correct

Standard answers take for granted that the Fed always knew that the rational expectations version of the natural rate hypothesis was correct

- Fluctuations in the natural rate (Parkin, Ireland)

Standard answers take for granted that the Fed always knew that the rational expectations version of the natural rate hypothesis was correct

- Fluctuations in the natural rate (Parkin, Ireland)
- Expectations traps (Chari, Christiano, Eichenbaum, Albanesi)

Standard answers take for granted that the Fed always knew that the rational expectations version of the natural rate hypothesis was correct

- Fluctuations in the natural rate (Parkin, Ireland)
- Expectations traps (Chari, Christiano, Eichenbaum, Albanesi)
- Imperfect estimates of the natural rate (Orphanides)

Standard answers take for granted that the Fed always knew that the rational expectations version of the natural rate hypothesis was correct

- Fluctuations in the natural rate (Parkin, Ireland)
- Expectations traps (Chari, Christiano, Eichenbaum, Albanesi)
- Imperfect estimates of the natural rate (Orphanides)
- Evil motives of Arthur Burns (DeLong)

Our explanation features learning

Our explanation features learning

- The Fed keeps several models on the table, including ones that don't incorporate the rational expectations version of the natural rate hypothesis ●.

Our explanation features learning

- The Fed keeps several models on the table, including ones that don’t incorporate the rational expectations version of the natural rate hypothesis ●.
- The Fed updates parameters of each model and its posterior probabilities over models

Our explanation features learning

- The Fed keeps several models on the table, including ones that don't incorporate the rational expectations version of the natural rate hypothesis ●.
- The Fed updates parameters of each model and its posterior probabilities over models
- The Fed uses dynamic programming to choose a first-period decision each period

Anticipated utility model (Kreps)

Anticipated utility model (Kreps)

- Using time t model, solve dynamic programming problem at t.

Anticipated utility model (Kreps)

- Using time t model, solve dynamic programming problem at t.
- Take time t decision

Anticipated utility model (Kreps)

- Using time t model, solve dynamic programming problem at t.
- Take time t decision
- Update model

Anticipated utility model (Kreps)

- Using time t model, solve dynamic programming problem at t.
- Take time t decision
- Update model
- Using time $t+1$ model, solve dynamic programming problem at $t+1$ and $\ldots$

Three models

- Samuelson-Solow model (a long-run tradeoff)

Three models

- Samuelson-Solow model (a long-run tradeoff)
- Solow-Tobin model (no ‘long-run’ tradeoff)

Three models

- Samuelson-Solow model (a long-run tradeoff)
- Solow-Tobin model (no ‘long-run’ tradeoff)
- Lucas model (no tradeoff)

Variables

- u_t unemployment

Variables

- u_t unemployment
- u_t^* natural rate

Variables

- u_t unemployment
- u_t^* natural rate
- y_t actual inflation

Variables

- u_t unemployment
- u_t^* natural rate
- y_t actual inflation
- x_t expected inflation

Three models

- Samuelson-Solow model

$$y_t = \gamma_0 + \gamma_1(L)y_{t-1} + \gamma_2(L)u_t + \eta_{1t}$$

Three models

- Samuelson-Solow model

$$y_t = \gamma_0 + \gamma_1(L)y_{t-1} + \gamma_2(L)u_t + \eta_{1t}$$

- Solow-Tobin model

$$\Delta y_t = \delta_1(L)\Delta y_{t-1} + \delta_2(L)(u_t - u^*_{t-1}) + \eta_{2t}$$

Three models

- Samuelson-Solow model

$$y_t = \gamma_0 + \gamma_1(L)y_{t-1} + \gamma_2(L)u_t + \eta_{1t}$$

- Solow-Tobin model

$$\Delta y_t = \delta_1(L)\Delta y_{t-1} + \delta_2(L)(u_t - u^*_{t-1}) + \eta_{2t}$$

- Lucas model

$$u_t - u^*_t = \phi_1(y_t - x_{t|t-1}) + \phi_2(L)(u_{t-1} - u^*_t) + \eta_{3t}$$

Three models

- Samuelson-Solow model

$$y_t = \gamma_0 + \gamma_1(L)y_{t-1} + \gamma_2(L)u_t + \eta_{1t}$$

- Solow-Tobin model

$$\Delta y_t = \delta_1(L)\Delta y_{t-1} + \delta_2(L)(u_t - u^*_{t-1}) + \eta_{2t}$$

- Lucas model

$$u_t - u^*_t = \phi_1(y_t - x_{t|t-1}) + \phi_2(L)(u_{t-1} - u^*_t) + \eta_{3t}$$

- η_{it} is iid $N(0, \sigma^2_i)$.

Form of each model

$$Y_t = X_t\theta + \nu_t$$

Priors

$$p(\theta, \sigma^2) = p(\theta|\sigma^2)p(\sigma^2)$$

The marginal prior $p(\sigma^2)$ makes the error variance an inverse gamma variate.

The conditional prior $p(\theta|\sigma^2)$ makes the regression parameters a normal random vector.

Posteriors

$$p(\theta|\sigma^2, Z^{t-1}) = N(\theta_{t-1}, \sigma^2 P_{t-1}^{-1}),$$

$$p(\sigma^2|Z^{t-1}) = IG(s_{t-1}, v_{t-1}),$$

Posteriors

$$p(\theta|\sigma^2, Z^{t-1}) = N(\theta_{t-1}, \sigma^2 P_{t-1}^{-1}),$$
$$p(\sigma^2|Z^{t-1}) = IG(s_{t-1}, v_{t-1}),$$

P_{t-1} is a precision matrix, s_{t-1} a scale parameter for the inverse-gamma density, and v_{t-1} counts degrees of freedom.

Posteriors

$$p(\theta|\sigma^2, Z^{t-1}) = N(\theta_{t-1}, \sigma^2 P_{t-1}^{-1}),$$
$$p(\sigma^2|Z^{t-1}) = IG(s_{t-1}, v_{t-1}),$$

P_{t-1} is a precision matrix, s_{t-1} a scale parameter for the inverse-gamma density, and v_{t-1} counts degrees of freedom. The estimate of σ^2 is just s_{t-1}/v_{t-1}.

Posteriors

After seeing outcomes at t, the central bank's updated beliefs are

$$p(\theta|\sigma^2, Z^t) = N(\theta_t, \sigma^2 P_t^{-1})$$

$$p(\sigma^2|Z^t) = IG(s_t, v_t)$$

Posteriors

After seeing outcomes at t, the central bank's updated beliefs are

$$p(\theta|\sigma^2, Z^t) = N(\theta_t, \sigma^2 P_t^{-1})$$
$$p(\sigma^2|Z^t) = IG(s_t, v_t)$$

where

$$P_t = P_{t-1} + X_t X_t',$$
$$\theta_t = P_t^{-1}(P_{t-1}\theta_{t-1} + X_t Y_t).$$
$$s_t = s_{t-1} + Y_t' Y_t + \theta_{t-1}' P_{t-1}\theta_{t-1} - \theta_t' P_t \theta_t,$$
$$v_t = v_{t-1} + 1.$$

Model probabilities

Let $\alpha_{i0} = p(M_i)$ represent the prior probability on model i.

Model probabilities

Let $\alpha_{i0} = p(M_i)$ represent the prior probability on model i. Bayes' theorem implies the posterior probability is

$$p(M_i|Y^t, X^t) \propto m_{it} \cdot p(M_i) \equiv w_{it},$$

where m_{it} is the marginalized likelihood function for model i at date t.

Model probabilities

Let $\alpha_{i0} = p(M_i)$ represent the prior probability on model i. Bayes' theorem implies the posterior probability is

$$p(M_i|Y^t, X^t) \propto m_{it} \cdot p(M_i) \equiv w_{it},$$

where m_{it} is the marginalized likelihood function for model i at date t. The conditional likelihood for model i through date t is defined via a prediction error decomposition as

$$l(Y^t, X^t, \theta, \sigma^2) = \prod_{s=1}^{t} p(Y_s|X_s, \theta, \sigma^2).$$

Model probabilities (2)

The marginalized likelihood is

$$m_{it} = \iint l(Y_i^t, X_i^t, \theta_i, \sigma_i^2) p(\theta_i, \sigma_i^2) d\theta_i d\sigma_i^2$$

Model probabilities (2)

The marginalized likelihood is

$$m_{it} = \iint l(Y_i^t, X_i^t, \theta_i, \sigma_i^2) p(\theta_i, \sigma_i^2) d\theta_i d\sigma_i^2$$

m_{it} is the normalizing constant in Bayes' theorem and can also be expressed as

$$m_{it} = \frac{l(Y_i^t, X_i^t, \theta_i, \sigma_i^2) p(\theta_i, \sigma_i^2)}{p(\theta_i, \sigma_i^2 | Z_i^t)}$$

Posterior model probabilities

$$\log w_{it+1} = \log w_{it} + \log p(Y_{it+1}|X_{it+1}, \theta_i, \sigma_i^2) - \log \frac{p(\theta_i, \sigma_i^2|Z_i^{t+1})}{p(\theta_i, \sigma_i^2|Z_i^t)}$$

Posterior model probabilities

$$\log w_{it+1} = \log w_{it} + \log p(Y_{it+1}|X_{it+1}, \theta_i, \sigma_i^2) -$$
$$\log \frac{p(\theta_i, \sigma_i^2 | Z_i^{t+1})}{p(\theta_i, \sigma_i^2 | Z_i^t)}$$

A renormalization enforces that the model weights sum to 1

$$\alpha_{it} = \frac{w_{it}}{w_{1t} + w_{2t} + w_{3t}}$$

Posterior model probabilities (2)

or

$$\alpha_{it} = [\exp R_{1i}(t) + \exp R_{2i}(t) + \exp R_{3i}(t)]^{-1}$$

where $R_{ji}(t) = (\log w_{jt} - \log w_{it})$ summarizes the weight of the evidence favoring model j relative to model i.

Posterior model probabilities (2)

or

$$\alpha_{it} = [\exp R_{1i}(t) + \exp R_{2i}(t) + \exp R_{3i}(t)]^{-1}$$

where $R_{ji}(t) = (\log w_{jt} - \log w_{it})$ summarizes the weight of the evidence favoring model j relative to model i.

State space models

$$S_{it+j} = A_i(t-1)S_{it+j-1} + B_i(t-1)x_{t+j|t-1} + C_i(t-1)\eta_{it+j} \quad (1)$$

where $(S_{it}, A_i(t-1), B_i(t-1), C_i(t-1))$ are the state vector and system arrays for model i at time t.

Grand model

$$\begin{bmatrix} S_{1t+j} \\ S_{2t+j} \\ S_{3t+j} \end{bmatrix} = \begin{bmatrix} A_1(t-1) & 0 & 0 \\ 0 & A_2(t-1) & 0 \\ 0 & 0 & A_3(t-1) \end{bmatrix} \begin{bmatrix} S_{1t+j-1} \\ S_{2t+j-1} \\ S_{3t+j-1} \end{bmatrix}$$

$$+ \begin{bmatrix} B_1(t-1) \\ B_2(t-1) \\ B_3(t-1) \end{bmatrix} x_{t+j|t-1}$$

$$+ \begin{bmatrix} C_1(t-1) & 0 & 0 \\ 0 & C_2(t-1) & 0 \\ 0 & 0 & C_3(t-1) \end{bmatrix} \begin{bmatrix} \eta_{1t} \\ \eta_{2t} \\ \eta_{3t} \end{bmatrix}$$

Grand model

$$\begin{bmatrix} S_{1t+j} \\ S_{2t+j} \\ S_{3t+j} \end{bmatrix} = \begin{bmatrix} A_1(t-1) & 0 & 0 \\ 0 & A_2(t-1) & 0 \\ 0 & 0 & A_3(t-1) \end{bmatrix} \begin{bmatrix} S_{1t+j-1} \\ S_{2t+j-1} \\ S_{3t+j-1} \end{bmatrix}$$

$$+ \begin{bmatrix} B_1(t-1) \\ B_2(t-1) \\ B_3(t-1) \end{bmatrix} x_{t+j|t-1}$$

$$+ \begin{bmatrix} C_1(t-1) & 0 & 0 \\ 0 & C_2(t-1) & 0 \\ 0 & 0 & C_3(t-1) \end{bmatrix} \begin{bmatrix} \eta_{1t} \\ \eta_{2t} \\ \eta_{3t} \end{bmatrix}$$

or in a more compact notation

$$S_{Et+j} = A_E(t-1)S_{Et+j-1} + B_E(t-1)x_{t+j|t-1} + C_E(t+j)\eta_t$$

Loss function

$$\mathcal{L}(M_i)(t) = E_t \sum_{j=0}^{\infty} \beta^j \big(S'_{it+j} M'_{s_i} Q M_{s_i} S_{it+j} + x'_{t+j|t-1} R x_{t+j|t-1} \big)$$

Grand loss function

$$\mathcal{L}_E(t) = \alpha_{1t}\mathcal{L}(M_1) + \alpha_{2t}\mathcal{L}(M_2) + \alpha_{3t}\mathcal{L}(M_3)$$

$$= E_t \sum_{i=1}^{3} \alpha_{it} \sum_{j=0}^{\infty} \beta^j (S'_{it+j} M'_{s_i} Q M_{s_i} S_{it+j}$$

$$+ x'_{t+j|t-1} R x_{t+j|t-1})$$

Grand loss function

$$\mathcal{L}_E(t) = \alpha_{1t}\mathcal{L}(M_1) + \alpha_{2t}\mathcal{L}(M_2) + \alpha_{3t}\mathcal{L}(M_3)$$

$$= E_t \sum_{i=1}^{3} \alpha_{it} \sum_{j=0}^{\infty} \beta^j (S'_{it+j} M'_{s_i} Q M_{s_i} S_{it+j}$$

$$+ x'_{t+j|t-1} R x_{t+j|t-1})$$

Grand loss function

$$\mathcal{L}_E(t) = \alpha_{1t}\mathcal{L}(M_1) + \alpha_{2t}\mathcal{L}(M_2) + \alpha_{3t}\mathcal{L}(M_3)$$

$$= E_t \sum_{i=1}^{3} \alpha_{it} \sum_{j=0}^{\infty} \beta^j (S'_{it+j} M'_{s_i} Q M_{s_i} S_{it+j}$$

$$+ x'_{t+j|t-1} R x_{t+j|t-1})$$

$$= E_t \sum_{j=0}^{\infty} \beta^j (S'_{E,t+j} Q_E S_{E,t+j} + x'_{t+j|t-1} R x_{t+j|t-1})$$

Grand loss function

$$\begin{aligned}\mathcal{L}_E(t) &= \alpha_{1t}\mathcal{L}(M_1) + \alpha_{2t}\mathcal{L}(M_2) + \alpha_{3t}\mathcal{L}(M_3) \\ &= E_t \sum_{i=1}^{3} \alpha_{it} \sum_{j=0}^{\infty} \beta^j (S'_{it+j} M'_{s_i} Q M_{s_i} S_{it+j} \\ &+ x'_{t+j|t-1} R x_{t+j|t-1}) \\ &= E_t \sum_{j=0}^{\infty} \beta^j (S'_{E,t+j} Q_E S_{E,t+j} + x'_{t+j|t-1} R x_{t+j|t-1})\end{aligned}$$

where

$$Q_E = \begin{bmatrix} \alpha_{1t} M'_{s_1} Q M_{s_1} & 0 & 0 \\ 0 & \alpha_{2t} M'_{s_2} Q M_{s_2} & 0 \\ 0 & 0 & \alpha_{3t} M'_{s_3} Q M_{s_3} \end{bmatrix}$$

Bellman equation

$$v_t(S_E) = \max_x \left\{ -S_E' Q_{Et} S_E - x' R x + \beta E v_t(S_E^*) \right\}$$

subject to

$$S_E^* = A_E(t-1) S_E + B_E(t-1) x + C_E(t-1) \eta$$

Bellman equation

$$v_t(S_E) = \max_x \left\{ -S_E' Q_{Et} S_E - x' R x + \beta E v_t(S_E^*) \right\}$$

subject to

$$S_E^* = A_E(t-1) S_E + B_E(t-1) x + C_E(t-1) \eta$$

Decision rule

$$x_{t|t-1} = -f_E(t-1) \cdot S_{Et-1} = -f_E(t-1)^1 S_{1t-1} - f_E^2(t-1) S_{2t-1} - f_E^3(t-1) S_{3t-1}$$

Decision rule

$$x_{t|t-1} = -f_E(t-1) \cdot S_{Et-1} = -f_E(t-1)^1 S_{1t-1} - f_E^2(t-1) S_{2t-1} - f_E^3(t-1) S_{3t-1}$$

If the composite model is ‘detectable’ and ‘stabilizable,’ then the policy rule f_E can be computed using standard algorithms

Data

- Inflation: log difference in chain weighted GDP deflator
- Unemployment: civilian unemployment rate
- quarterly, seasonally adjusted, 1948:Q1–2002:Q4
- $u_t^* = u_{t-1}^* + .075(u_t - u_{t-1}^*)$

Parameters

- $\beta = 1.04^{.25}$
- $\alpha_{10} = .98, \alpha_{20} = \alpha_{30} = .01$
- $\lambda = 16$ (equal weight on u, y)
- Training sample: first 12 years

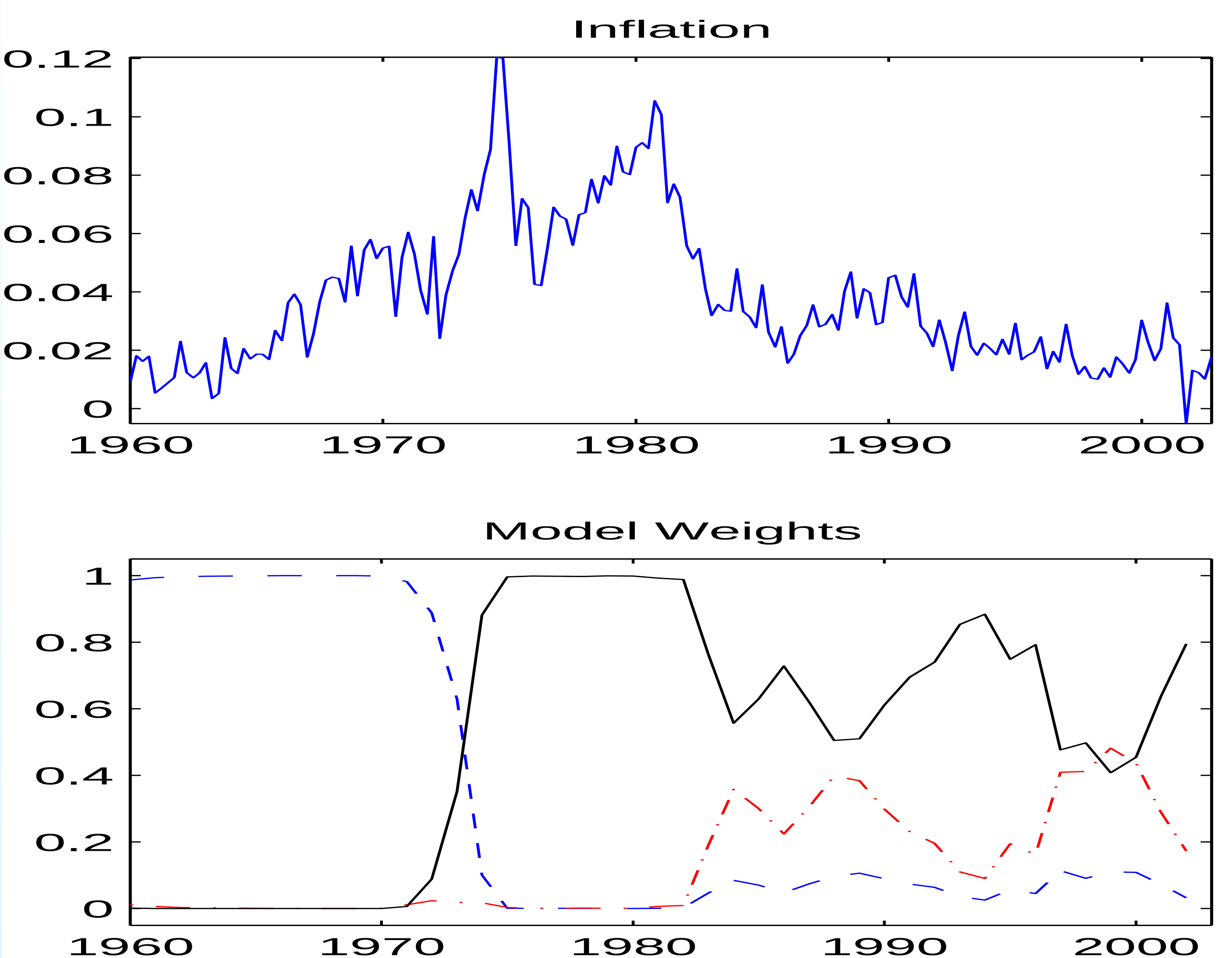
Lag orders

	Inflation	Unemployment
Samuelson-Solow	$\gamma_1 : 4$	
Solow-Tobin	$\delta_1 : 3$	
Lucas-Sargent	$\phi_1 : 0$	

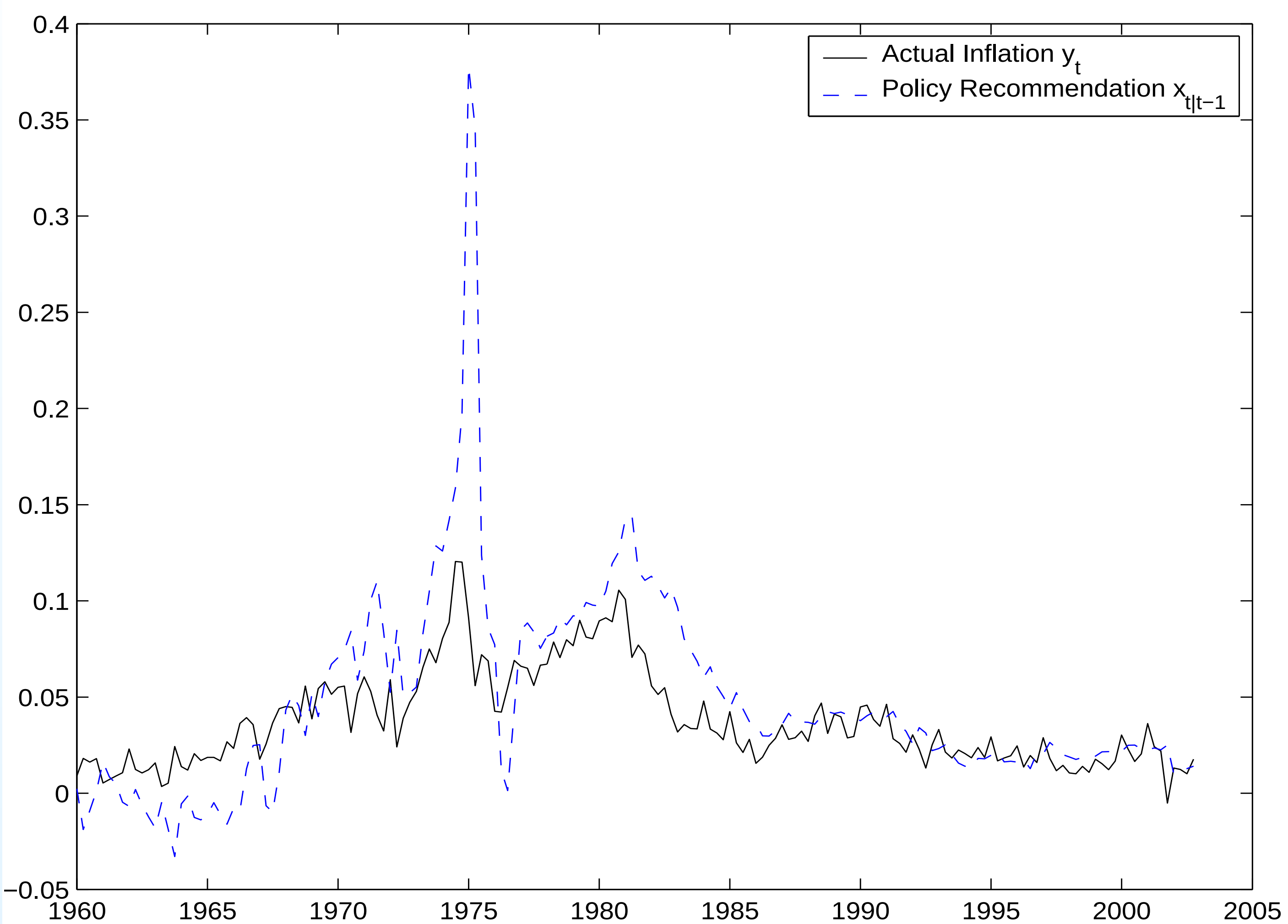
Lag orders

	Inflation	Unemployment
Samuelson-Solow	$\gamma_1 : 4$	$\gamma_2 : 2$
Solow-Tobin	$\delta_1 : 3$	$\delta_2 : 2$
Lucas-Sargent	$\phi_1 : 0$	$\phi_2 : 2$

Inflation and α_{it}'s

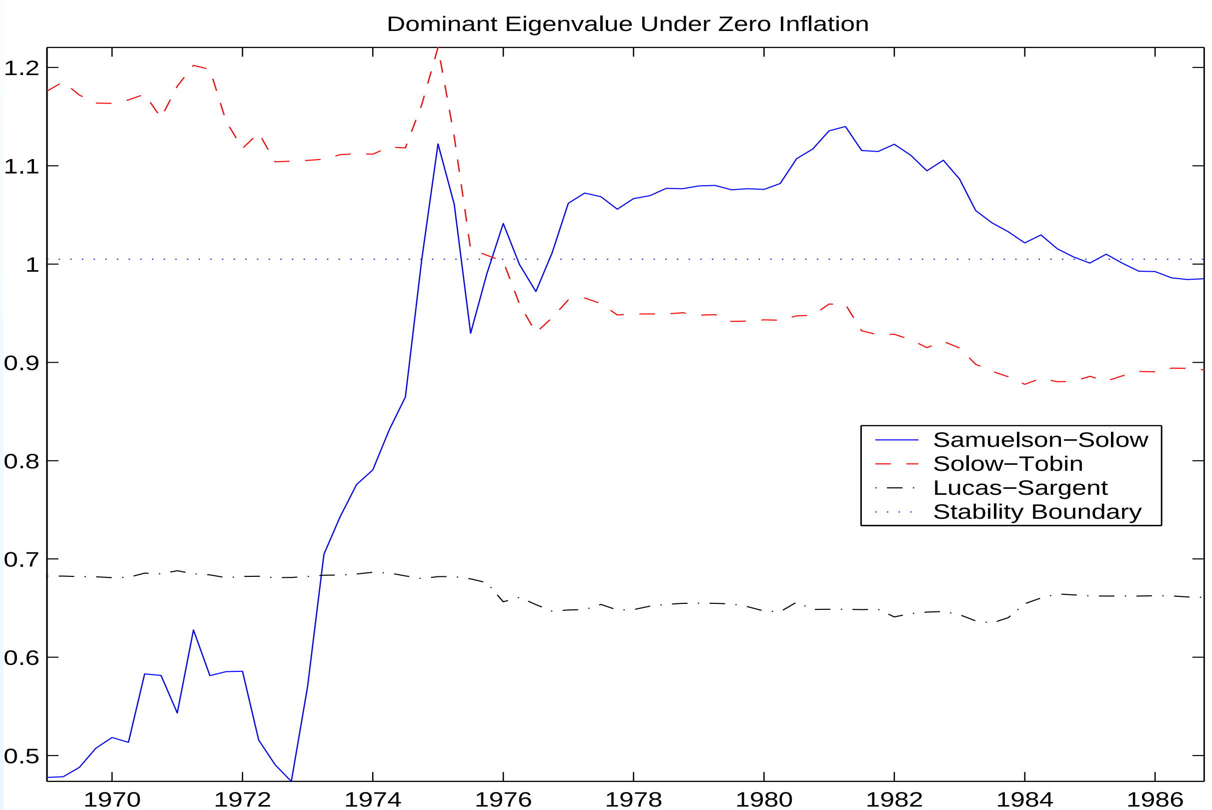
- 

Inflation and Optimal Policy

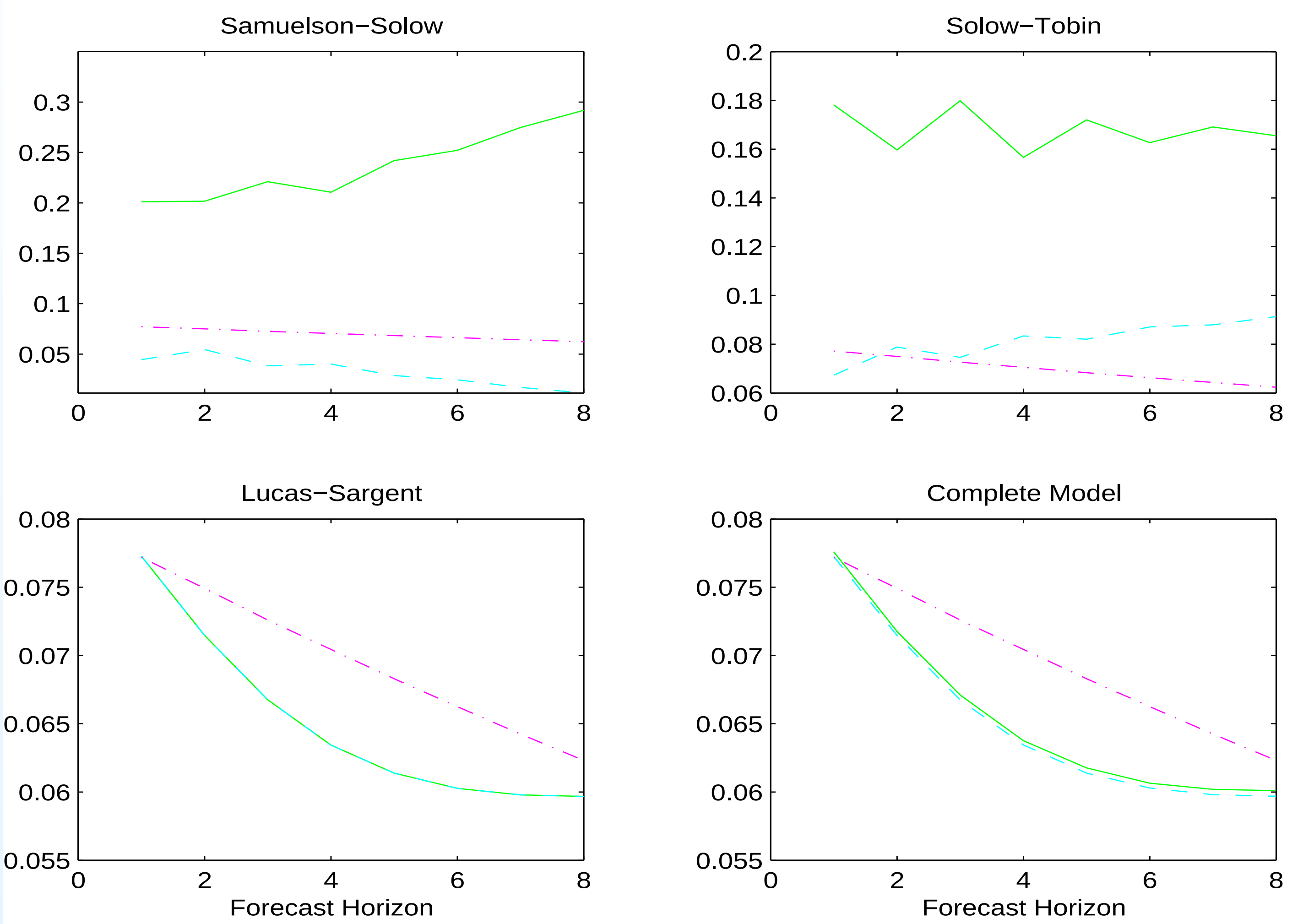


Expected Loss from a Zero Inflation Policy

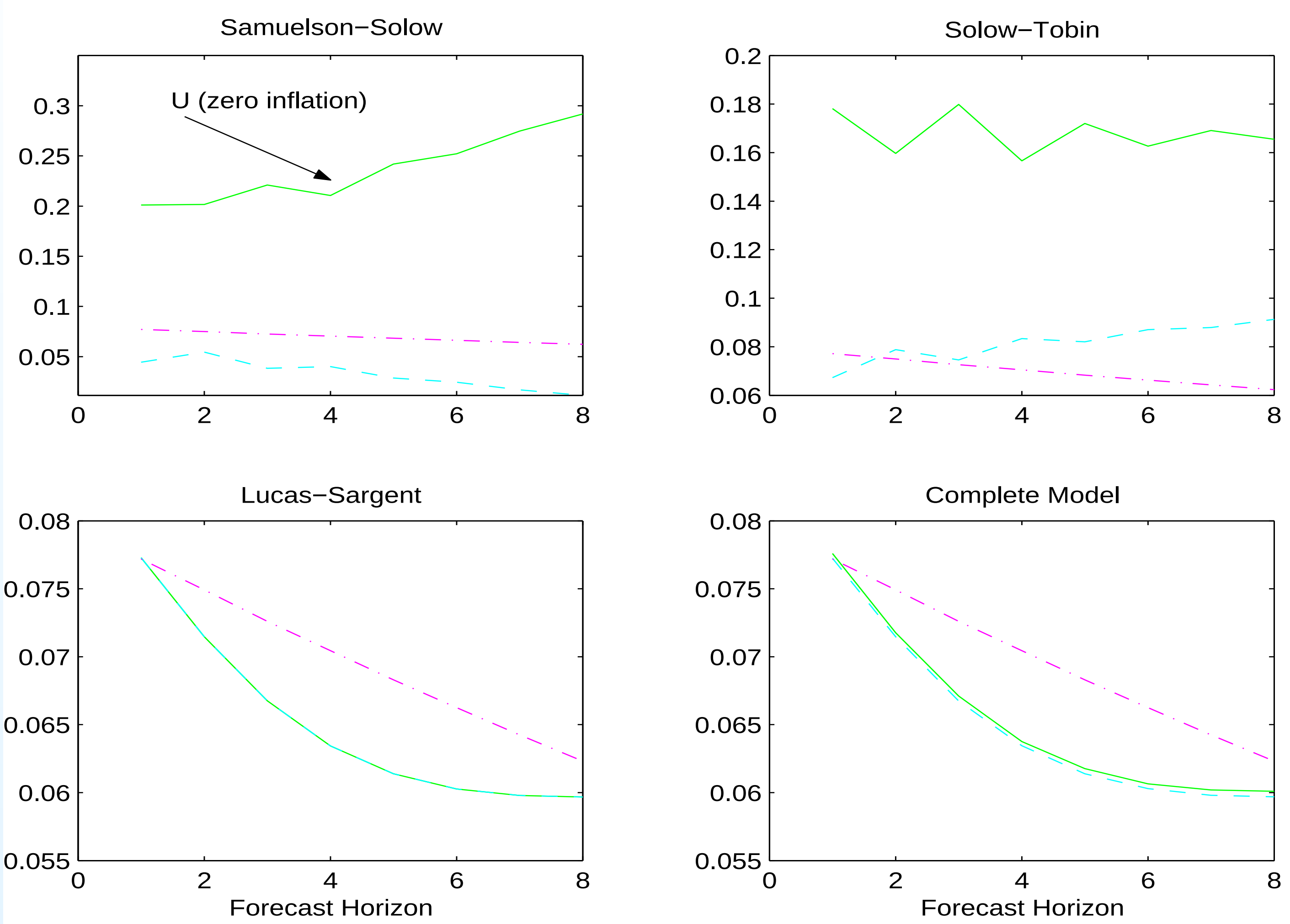
Dominant Eigenvalue Under Zero Inflation



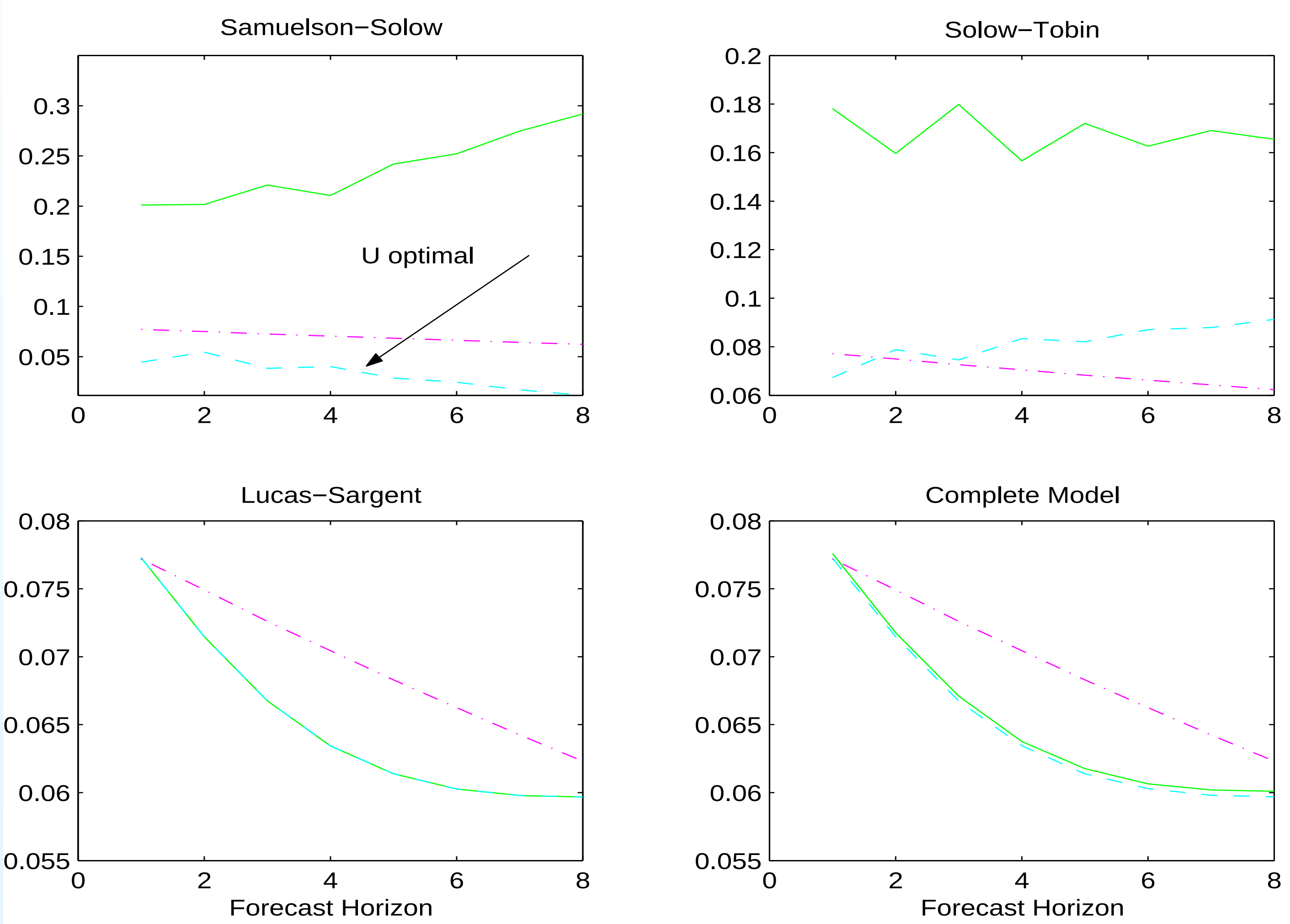
Optimal Policy v. Zero Inflation, 1975.Q4



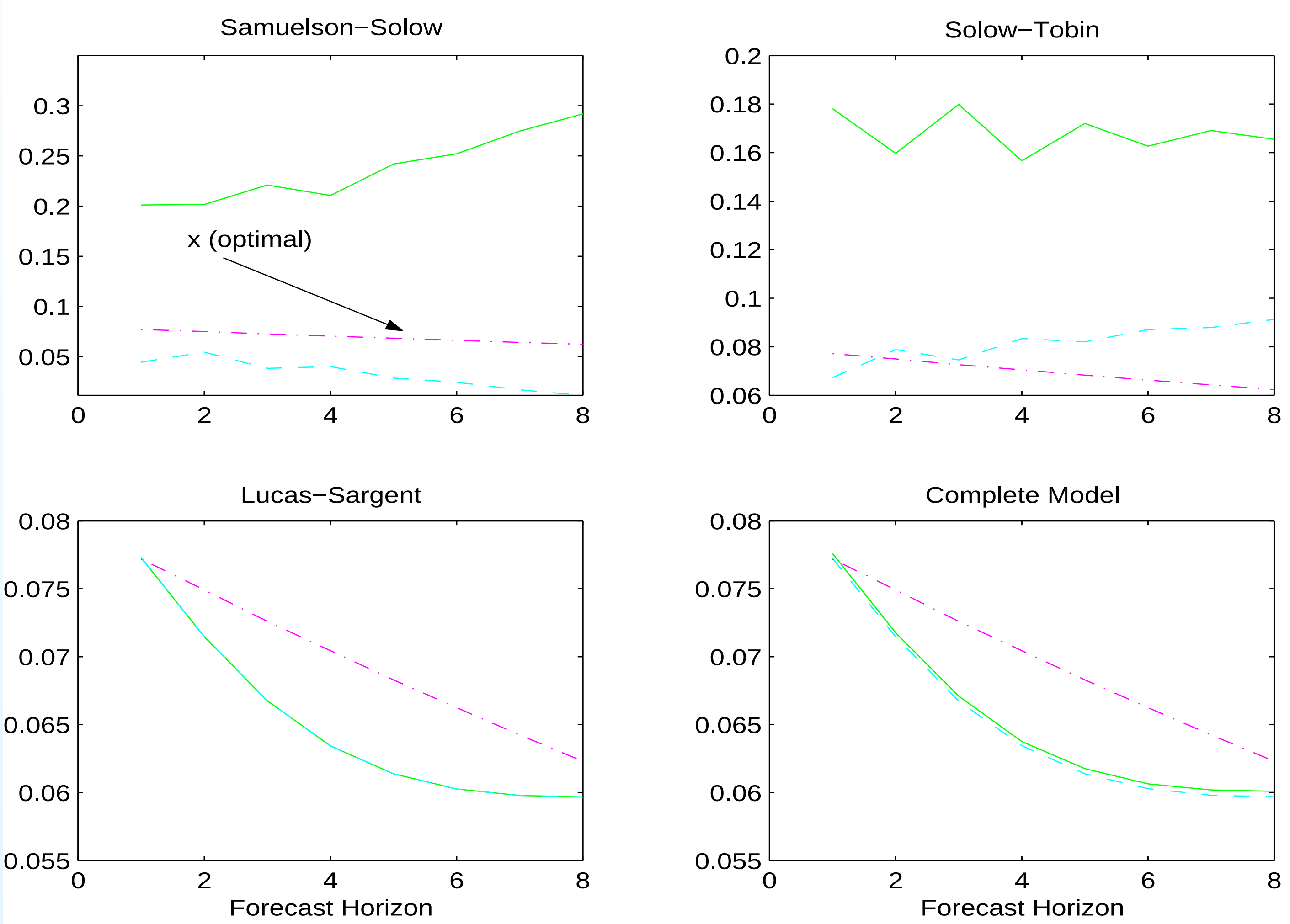
Optimal Policy v. Zero Inflation, 1975.Q4



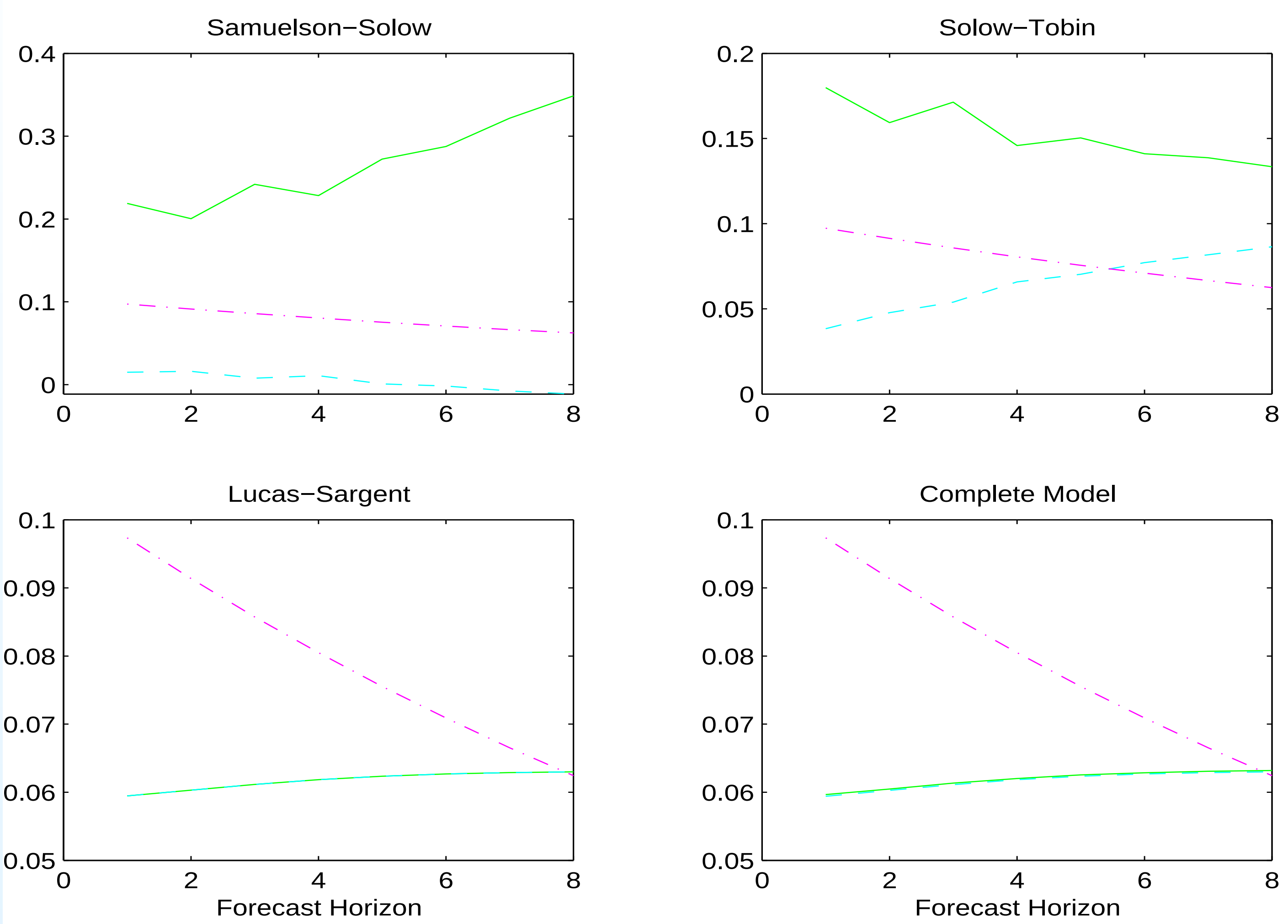
Optimal Policy v. Zero Inflation, 1975.Q4



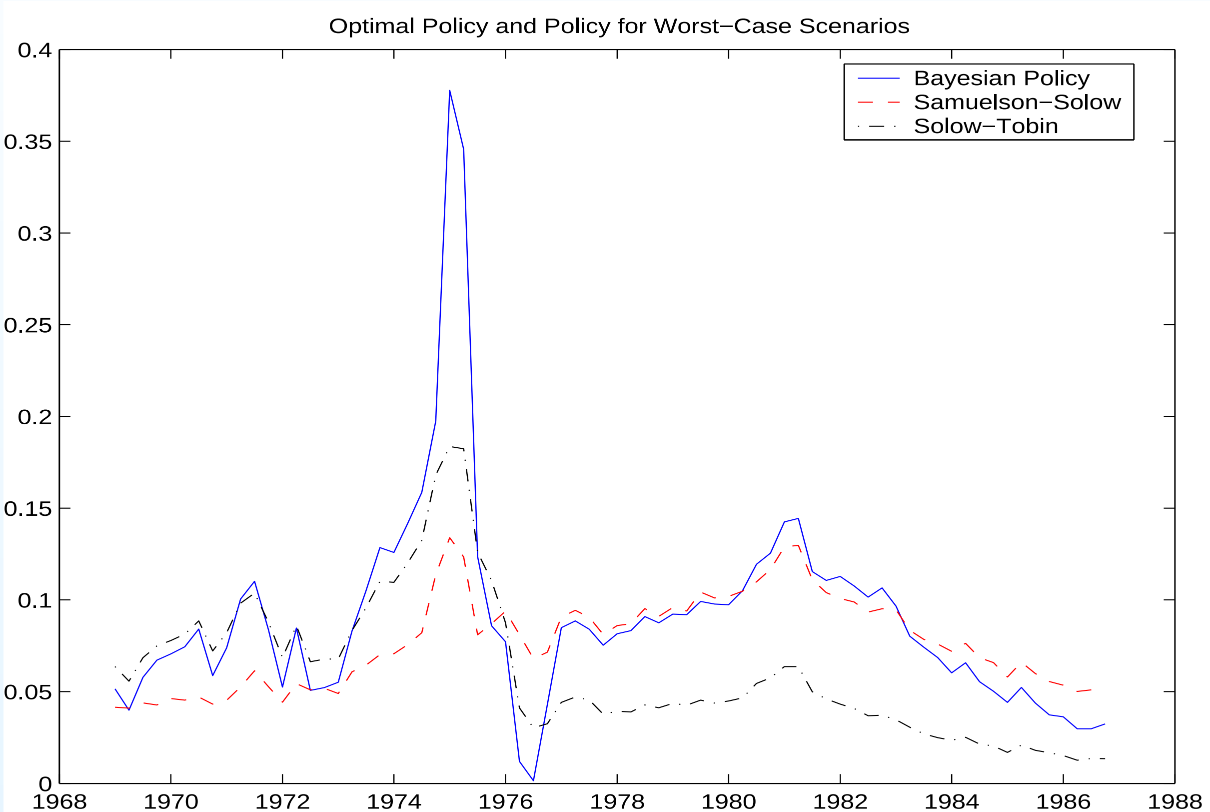
Optimal Policy v. Zero Inflation, 1975.Q4



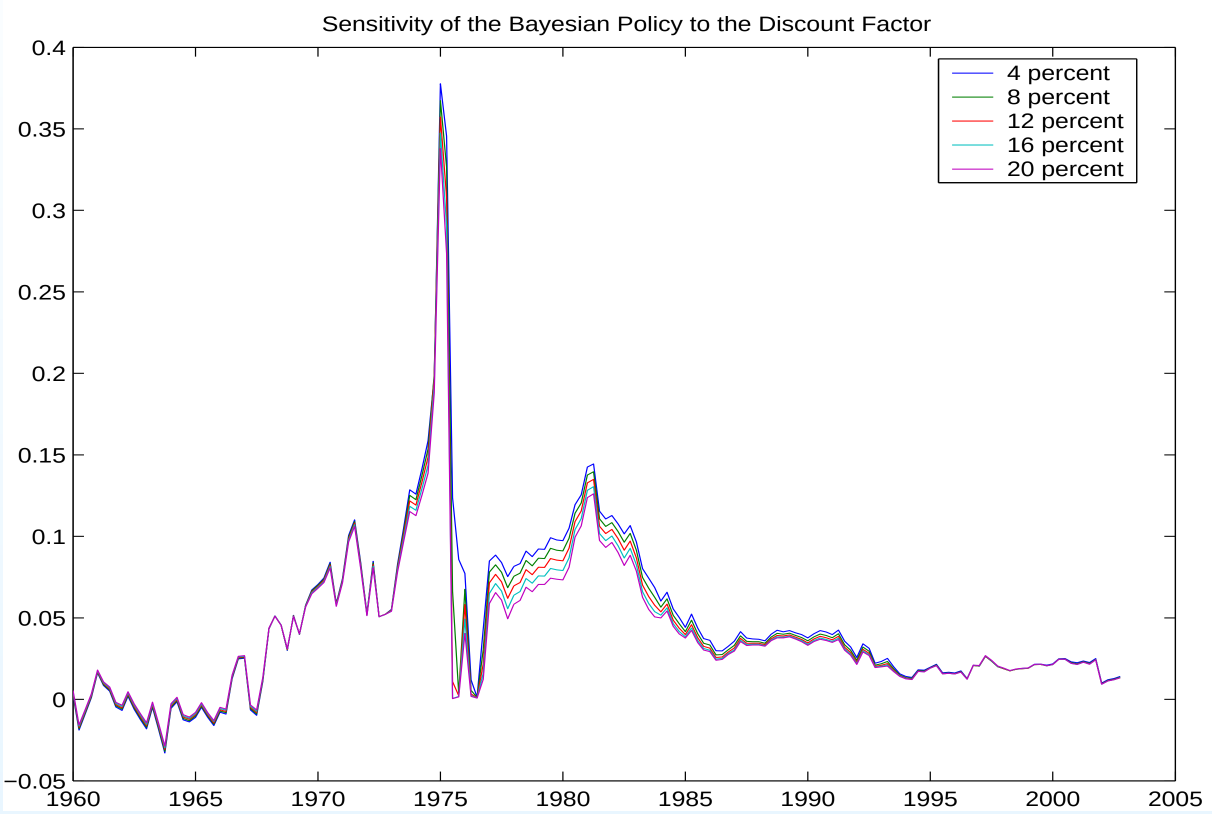
Optimal Policy v. Zero Inflation, 1979.Q4



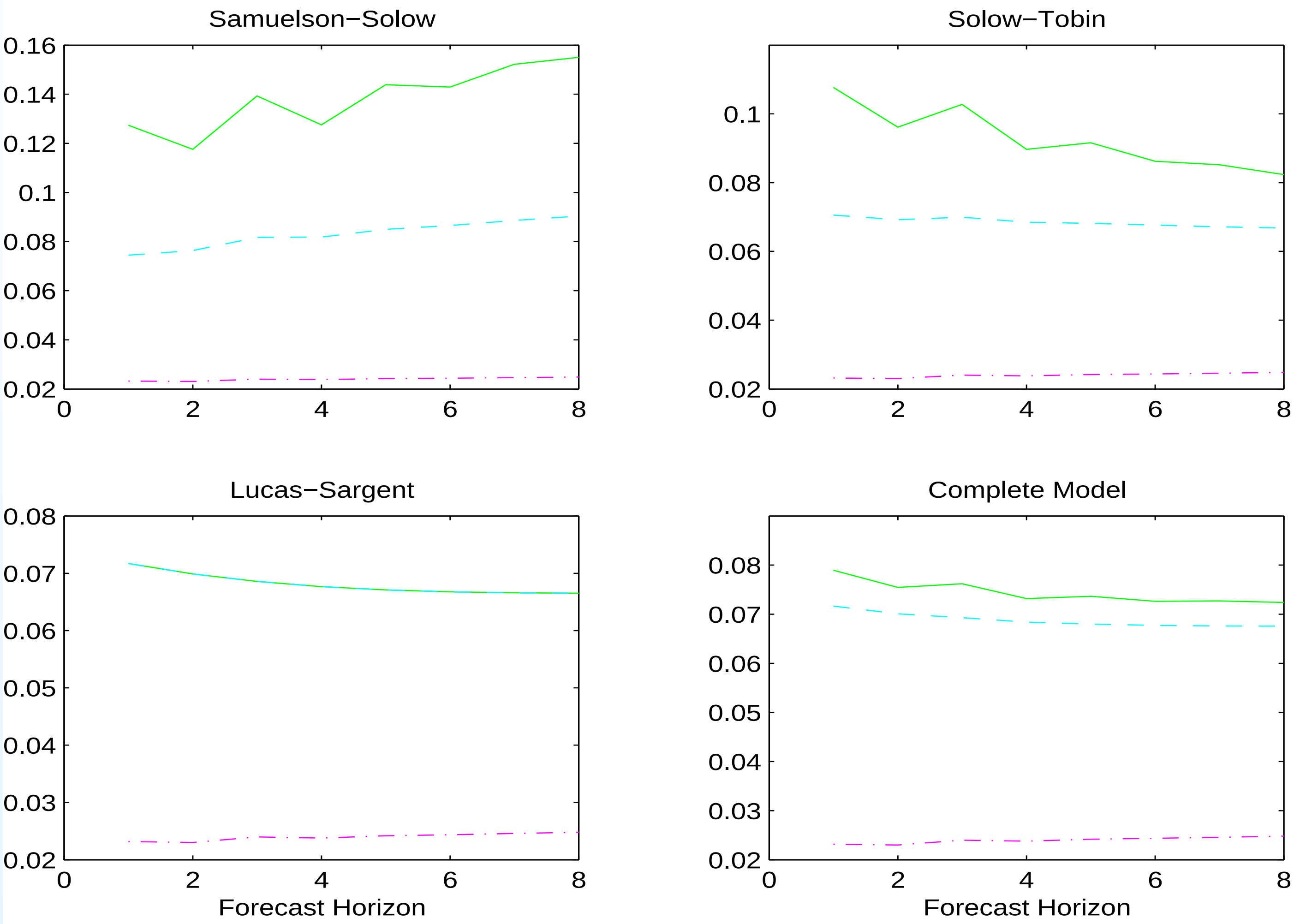
Optimal Policy and Policy for Worst-Case Scenarios



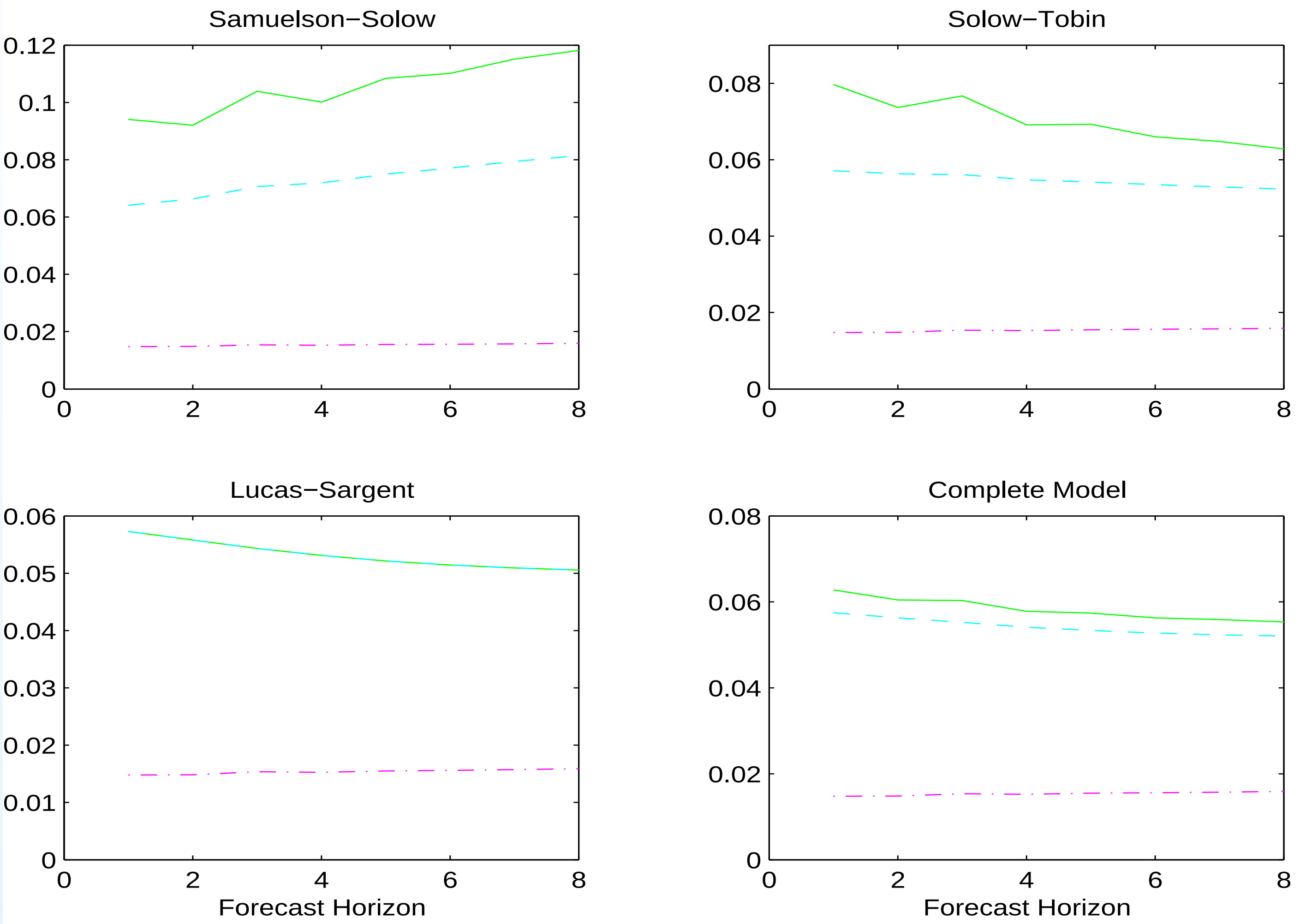
Sensitivity of Bayesian Policy to the Discount Rate



Optimal Policy v. Zero Inflation, 1992.Q4



Optimal Policy v. Zero Inflation, 2002.Q4



1978

Okun and Perry (1978) summarize things as follows:

> "Thus, the mainline model and its empirical findings reaffirm that there is a slow-growth, high unemployment cure for inflation, but that it is an extremely expensive one. ...Using one of Perry's successful equations as an example, an extra percentage point of unemployment would lower the inflation rate by only about 0.3 percentage point after one year and by 0.7 percentage point if maintained for three years. That extra point of unemployment would cost over a million jobs and some $60 billion of real production each year."

1978 again

Okun and Perry also summarize Perry's reasons for rejecting Feller's guesses of much lower costs in terms of unemployment that could be attained through a credible disinflationary policy:

> Perry "believes that much of the [inflation] inertia is backward-looking rather than forward looking, and so is not susceptible to even convincing demonstrations that demand will be restrained in the future. [Perry's] own empirical evidence shows that wage developments are better explained in terms of the recent past history of wages and prices than on any assumption that people are predicting the future course of wages and prices in a way that differs from the past." (page 6)

More 1978 beliefs

Perry (1978, pp. 50-51) elaborates forcefully on his argument against an expectational interpretation of Phillips curve dynamics. Okun (1978a, p.284) says that “recession will slow inflation, but only at the absurd cost in production of roughly $200 billion per point.” Okun (1978b) lists the models on which this estimate is based. Many distinguished economists are represented on the list. At that time, $200 billion amounted to roughly 10 percent of GDP. Inflation averaged 7.4 percent from 1974 to 1979, and extrapolating to zero inflation implies a total cost of almost three-quarters of a year’s GDP.