The Output Cost of Disinflation in Traditional and Vector Autoregressive Models
Author(s): Robert J. Gordon, Stephen R. King and  Franco Modigliani
Source: *Brookings Papers on Economic Activity,* Vol. 1982, No. 1 (1982), pp. 205-244
Published by: Brookings Institution Press
Stable URL: https://www.jstor.org/stable/2534320
Accessed: 23-04-2020 23:36 UTC

REFERENCES
Linked references are available on JSTOR for this article:
https://www.jstor.org/stable/2534320?seq=1&cid=pdf-reference#references_tab_contents
You may need to log in to JSTOR to access the linked references.

ROBERT J. GORDON
*Northwestern University*

STEPHEN R. KING
*Northwestern University*

# The Output Cost of Disinflation in Traditional and Vector Autoregressive Models

THE SPEED of adjustment of the aggregate price level to demand and supply shocks has long been a leading topic of controversy in macroeconomics. Among the many issues requiring for their resolution solid empirical evidence on the dynamics of price adjustment is the prediction of the output loss that would accompany a strategy of monetary disinflation. Four years ago Arthur M. Okun surveyed a variety of econometric evidence and reached the pessimistic conclusion that the inflation process in the postwar United States is so inertia prone that the cumulative sacrifice of 10 percent of a year's GNP would be required to achieve a permanent 1 percentage point reduction in the inflation rate.[1]

This paper compares the dynamic response patterns of prices and output that emerge from two quite different approaches to time-series econometrics, the traditional structural framework imbedded in most econometric models, and the more recent nonstructural or atheoretical

1. Arthur M. Okun, "Efficient Disinflationary Policies," *American Economic Review,* vol. 68 (May 1978, *Papers and Proceedings, 1977*), pp. 348–52.

vector autoregressive (VAR) technique. Both approaches reach conclusions by imposing restrictions of different types; by assessing the validity of these restrictions, we are able to compare the merits of each methodology. Of equal importance are new estimates of the speed of price adjustment in the postwar United States, which we summarize in a single number called the sacrifice ratio that measures the output loss required to eliminate permanently one point of inflation. By introducing several channels of monetary influence on the inflation process that are often overlooked, we conclude that the sacrifice ratio is roughly half that suggested by Okun's survey.

Although they are often regarded as radically different, both the traditional and VAR approaches to time-series econometrics essentially carry out the same task of allocating zero restrictions in the face of scarce degrees of freedom. With only 140 quarterly observations available in the postwar U.S. national accounts data for 1947–81, an econometric model containing sixteen endogenous and exogenous variables would have only four degrees of freedom remaining if each variable were entered with eight lagged values on the right-hand side of each equation.[2] The traditional approach uses theory to exclude all but a few variables from each equation—for instance, the investment tax credit matters for investment but not for wages—while price control dummies and energy prices matter for prices but not for consumption, and so on. This method of imposing zero restrictions allows econometric models to become very large and, if necessary, to contain more variables than there are sample observations available.

In contrast, the typical small-scale VAR model treats all variables symmetrically by including each on the right-hand side of every equation and by allowing each explanatory variable to enter with the same number of lagged values. This symmetry forces investigators to limit the total number of variables in the model to an arbitrary subset believed to be important for the economy as a whole (interest rate, money, price level, output) and to exclude variables that the traditional approach typically includes in individual equations (investment tax credit, control dummies, energy prices).[3]

2. With eight lagged values, the first observation of each equation would be 1949:1, leaving 132 observations in the sample period.

3. The current popularity of VAR models attests to the influence of two papers by Christopher Sims, "Macroeconomics and Reality," *Econometrica,* vol. 48 (January 1980), pp. 1–48, and "Comparison of Interwar and Postwar Business Cycles: Monetarism

Christopher Sims has argued convincingly that many of the zero restrictions embodied in traditional models are "incredible," particularly because any lagged variable may influence the formation of expectations. Our paper makes the reverse criticism that the zero restrictions embodied in VAR models are equally dubious because the pursuit of symmetry has usually led investigators to exclude explanatory variables that other research demonstrates to be highly significant statistically in some equations, and to have a strong theoretical presumption of relevance. Our preferred hybrid strategy for model specification uses the VAR approach to evaluate conventional restrictions and exogeneity assumptions, but then includes a second step that "edits" insignificant variables and lag lengths, as well as nominal variables from equations explaining relative price variables, to obtain a model of tractable size that yields plausible relations in long-run simulations.

Our use of alternative models to calculate sacrifice ratios for hypothetical *future* policy regimes is subject to the Lucas critique that parameters estimated from sample-period values may not be invariant to arbitrary shifts in policy. Most papers using the VAR technique have avoided this critique by constructing multivariate exogeneity and causality tests for small innovations to money or other variables assumed to occur within the historical sample period. We defend our excursion into the future against the Lucas critique by pointing to the stability of parameters in our basic inflation equation over a historical sample period during which the response of monetary policy to output and inflation underwent significant changes.

## The VAR Methodology

### ESTIMATION

The VAR methodology begins with the concept of a covariance-stationary time series, one that has a mean and an autocovariance at all

Reconsidered," *American Economic Review,* vol. 70 (May 1980 *Papers and Proceedings, 1979*), pp. 250–57. Some of the methodology was developed in Robert B. Litterman, "Techniques of Forecasting Using Vector Autoregressions," Working Paper 115 (Federal Reserve Bank of Minneapolis, 1979). Recent applications of the technique include Stanley Fischer, "Relative Shocks, Relative Price Variability, and Inflation," *BPEA, 2:1981,* pp. 381–431, and Benjamin Friedman, "The Roles of Money and Credit in Macroeconomic Analysis," Working Paper 831 (National Bureau of Economic Research, 1981).

lags that are constant through time.[4] By Wold's theorem any such time-series process, say $x_t$, can be decomposed into two components. The first, $\eta_t$, is linearly deterministic, that is, exactly predictable given a linear combination of its own past values; the second is a moving average, possibly of infinite length, of white noise errors, $\epsilon_t$:[5]

$$(1) \qquad x_t = \eta_t + A(L)\epsilon_t, \qquad E(\epsilon_t) \quad = 0$$

$$E(\epsilon_t \epsilon_{t-k}) = \begin{cases} \Omega & k = 0 \\ 0 & k \neq 0, \end{cases}$$

where $A(L)$ is a polynomial in the lag operator.

When the polynomial $A(L)$ is invertible,[6] an autoregressive representation of equation 1 exists and can be written as

$$(2) \qquad A(L)^{-1}x_t = A(L)^{-1}\eta_t + \epsilon_t.$$

By moving the lagged $x$'s to the right-hand side of the equation and combining them with the $\eta$'s, which, by definition, are linear functions of lagged $x$'s, we obtain the system of equations,

$$(3) \qquad x_t = B(L)x_t + \epsilon_t = \sum_{j=1}^{N} B_j L^j x_t + \epsilon_t.$$

In general, $N$, the lag length of the autoregressive representation in 3, will be infinite, but in practice it is generally truncated to some number that is both small enough to be computationally feasible and large enough to ensure that the equation residuals are approximately white noise. In

---

4. Covariance stationarity is not an innocuous assumption, but it can often be approximated for macroeconomic time series by defining variables as first differences.

5. White noise errors, like covariance stationary series, have constant autocovariances, but in addition have all covariances identically zero. That is, there are no systematic components that would enable a white noise process to be predicted from its own past.

6. Invertibility of $A(L)$ rules out cases in which $x_t$ depends to a greater extent on past innovations than on current ones. For example, if equation 1 were univariate, $\eta_t = 0$, and $A(L) = 1 - aL$, that equation would be $x_t = \epsilon_t - a\epsilon_{t-1}$. Successive substitutions to eliminate the lagged error terms would yield

$$x_t + ax_{t-1} + a^2 x_{t-2} + \cdots + a^m x_{t-m} = \epsilon_t - a^{m+1}\epsilon_{t-m-1}.$$

If $a$ is greater than 1, the last term does not vanish as $m$ increases, so no autoregressive representation exists. In this case, the requirement of invertibility for the polynomial $A(L) = 1 - aL$ is that $a$ is less than 1 in absolute value.

this case, 3 is the basic form of a vector autoregression in which each regressor $x_{it}$, an element of the vector $x_t$, is a linear function of its own lagged values, the lagged values of all other regressors in the system, and a white noise error term. If there are $M$ time-series variables in the model, then the coefficient matrix $B_j$ is of dimension $M$ by $NM$. As a consequence, every variable in the model is treated as being endogenous, and each has two components—its best linear predictor given information available one period previously, and its linearly unpredictable "innovation."

An example of the general form of 3 can be seen in a hypothetical VAR model containing only two variables, growth of the money supply, $m_t$, and $p_t$, the GNP deflator:[7]

(4a)                    $$m_t = b_{mm} m_{t-1} + b_{mp} p_{t-1} + \epsilon_{mt}$$

(4b)                    $$p_t = b_{pm} m_{t-1} + b_{pp} p_{t-1} + \epsilon_{pt}.$$

Here each variable is explained by one lag ($N = 1$) of each of the two ($M = 2$) variables in the model and an error or "innovation" term ($\epsilon_t$) that represents that part of the dependent variable not predictable from knowledge of lagged values of the regressors. Since we have two equations and one lagged value, the coefficient matrix $B$ is of dimension $2 \times 2$.

Equation 3 and the example (4a and 4b) take the form of the multivariate regression model, and the presence of identical sets of regressors for each of the $M$ equations ensures that the coefficients may be estimated consistently by single-equation least squares.[8] If it is further assumed that the innovations, $\epsilon_t$, are not only white noise but are also normally distributed, then the estimates of the $B_j$ coefficients are asymptotically efficient.

The testing of restrictions in a VAR is quite different from standard econometric methodology because it involves considering the impact of a given restriction on the model as a whole, rather than on each individual equation.[9] For instance, the test of truncation restrictions has generally

7. Throughout this paper lowercase variables denote rates of growth; uppercase denote levels.

8. Peter Schmidt, *Econometrics* (Marcel Dekker, 1976), pp. 78–80.

9. Tests of restrictions on the model can be carried out by comparing the determinants of the restricted and unrestricted covariance matrices of the equation errors. The test

been to test the joint significance of longer lags (such as eight versus four quarters) on all variables in all equations. Yet this procedure may reject longer lags that are unimportant in all equations except one, the one in which the lags may have a significant explanatory role. This is an example of how substantive economic issues become intertwined with restrictions that are said to be introduced simply to reduce complexity.

SIMULATIONS

All simulations calculated in VAR studies must grapple with the treatment of contemporaneous correlation among innovations. In conventional model building this issue is often suppressed by arbitrary restrictions that constrain the contemporaneous correlation between two variables to be unidirectional. This occurs, for instance, in models in which the money supply is treated as exogenous, and current money changes are included in an equation for price changes. In the VAR framework both prices and money are assumed to be endogenous, and because contemporary right-hand variables are omitted at the estimation stage, any contemporaneous correlation shows up as a correlation between the current innovations in the price and money equations.

Simulations of the effect of an exogenous shock require that some assumption be made about the causal ordering of the relation. Investi-

---

statistic $a$ can be computed as

$$a = (T - k)(\log |\Omega^R| - \log |\Omega^U|),$$

where $T$ is the number of observations, $k$ is the number of estimated parameters in each equation, and $|\Omega^R|$ and $|\Omega^U|$ denote, respectively, the determinants of the contemporaneous covariance matrix of the residuals of the restricted and unrestricted models. This statistic $a$ is distributed as $\chi^2$ with $r$ degrees of freedom, where $r$ is the number of restrictions imposed. If the $\Omega$ matrices are diagonal (implying that residuals are mutually uncorrelated across equations) the relevant determinants are simply the product of the residual sums of squares from each equation and the statistic $a$ clearly interpretable as the deterioration in fit caused by imposing the restrictions. If there were only one equation the statistic would reduce to approximately

$$\frac{(SSR^R - SSR^U)}{SSR^U/(T - k)},$$

where $SSR$ denotes the sum of squared residuals. This statistic is easily seen to be $r$ multiplied by the conventional $F$-statistic for testing restrictions in a single regression-equation.

gators can avoid an arbitrary choice about causal ordering only if they have a single-equation model, or if they are fortunate enough to find that the innovations in each equation, for instance, $\epsilon_{mt}$ and $\epsilon_{pt}$ in the example 4a, 4b are contemporaneously uncorrelated. In this lucky case, the estimated equations 4a and 4b can be inverted to compute the moving-average response of $p_t$ to current and past innovations,

$$(5) \qquad p_t = \epsilon_{pt} + b_{pp}\epsilon_{p,t-1} + (b_{pp}^2 + b_{pm}b_{mp})\epsilon_{p,t-2} + \ldots$$

$$+ \; b_{pm}\epsilon_{m,t-1} + (b_{pm}b_{mm} + b_{pp}b_{pm})\epsilon_{m,t-2} + \ldots,$$

and a symmetric response for $m_t$. In 5 a monetary innovation in period $t$ has no effect on prices until period $t + 1$, and vice versa for the effect of a price innovation on money. More generally, the estimated system of equations given by 3 can be inverted to compute $x_t$ as a moving average of past errors:

$$(6) \qquad\qquad x_t = (I - B(L))^{-1}\epsilon_t.$$

If, however, the innovation processes are contemporaneously correlated, investigators must decide how to treat this correlation. In our two-equation example, there are two obvious alternatives. First, the error in the money equation can be decomposed into a portion explained by the price innovation and a remaining independent portion, $u_{mt}$:

$$(7) \qquad\qquad \epsilon_{mt} = c_{mp}\epsilon_{pt} + u_{mt}; \qquad \epsilon_{pt} = u_{pt},$$

where $c_{mp}$ is the estimated coefficient in a regression of $\epsilon_{mt}$ on $\epsilon_{pt}$. The second alternative is to assume that the price error can be decomposed in the opposite direction:[10]

$$(8) \qquad\qquad \epsilon_{pt} = c_{pm}\epsilon_{mt} + u_{pt}; \qquad \epsilon_{mt} = u_{mt}.$$

Now consider introducing a shock, $s_m$, into the money equation equal to one sample-period standard deviation of the error, $\epsilon_m$, and comparing this event with another hypothetical situation in which no such shock occurs. The calculated effect of this on prices in the initial period would

10. Note that these two alternatives, and the third choice discussed below, do not exhaust the plausible assumptions about causality between contemporaneous errors. It would also be possible to assume that each error helps to explain the others. Then, however, regression techniques could not be used, and the size of each error's effects on the others would have to be known a priori.

be $\Delta p_t = 0$ under the alternative of 7 and in the second period would be $b_{pm}s_m$. In contrast, if the alternative of 8 were used, the initial-period response of prices would be $c_{pm}s_m$, and the second period response would be $(b_{pm} + b_{pp}c_{pm})s_m$. Thus it is likely that the simulation of a monetary disinflation using 8 would yield a larger and faster dynamic response of prices than an alternative simulation using 7.

At first glance it might seem preferable to avoid the choice between 7 and 8 by ignoring the contemporaneous correlation, that is, by setting both $c_{mp}$ and $c_{pm}$ equal to zero even though they are known to be nonzero.[11] This third choice would be tantamount to the selection of 7 for the simulation of a monetary shock, since the price responses in the first two periods would be, respectively, zero and $b_{pm}s_m$. And the use of the same criterion for the simulation of the effects of a price innovation would lead investigators into an inconsistency, since in this case they would have switched in midstream from 7 to 8. In short, the third choice is even more arbitrary than the first two. It is both inconsistent and involves throwing out known information.[12]

The assumption about causal ordering of contemporaneous errors in a VAR system amounts to a decision about admitting current variables into the estimating equation. To see this, return to the general VAR model in 3 and decompose each error term, $\epsilon_{it}$, into a part explained by the other innovations, $\epsilon_{jt}$, and a remaining component that is orthogonal to them, $u_t$.

Following the analysis given above for the two-variable case, we assume that if $\epsilon_{jt}$ affects $\epsilon_{it}$, there is no reverse causality. We order the variables so that a given error affects only errors that are lower in the

---

11. In this example, $c_{mp}$ and $c_{pm}$ are the regression coefficients from equations 7 and 8; hence,

$$c_{mp} = \sum_{t=1}^{T} \epsilon_{mt}\epsilon_{pt} / \sum \epsilon_{pt}^2 = \omega_{mp}/\omega_{pp},$$

where $\omega_{ij}$ is the $i, j$th element of $\Omega$.

In general, the $C$ matrix can be calculated recursively from the identity $(I - C)\epsilon_t = u_t$, and hence the identity

$$(I - C)^{-1}uu'(I - C)^{-1'} = \Omega.$$

12. In general, any linear combination of $c_{pm}$ and $c_{mp}$ would be acceptable, since equations 7 and 8 are both unidentified, but in looking at the extremes we are able to examine the full effect of the ordering assumption on the properties of the system.

list; that is, $\epsilon_{jt}$ affects $\epsilon_{it}$ only if $j < i$. This ordering is called a triangularization of the system. In matrix notation we can write a set of $M$ regression equations analogous to 7 and 8:

$$(9) \qquad\qquad\qquad \epsilon_t = C\epsilon_t + u_t,$$

where $C$ is a lower triangular $M \times M$ matrix with zeros on the diagonal, and whose $i, j$th element is the regression coefficient of $\epsilon_i$ on $\epsilon_j$ for $j < i$. Since the $\epsilon_t$ vector is orthogonal to all of the regressors in equation 3, the $B$ and $C$ coefficients could also be obtained by fitting the set of regressions,

$$(10) \qquad\qquad\qquad x_t = \sum_{j=1}^{N} B_j L^j x_t + C\epsilon_t + u_t,$$

where each equation except the first includes in the list of regressors the residuals from each previous regression. It is easy to show that identical residuals, $u_t$, to those in 10 will be obtained from an alternative set of regressions that directly include, in all equations except the first, the current values of the dependent variables from each previous equation,

$$(11) \qquad\qquad\qquad x_t = \sum_{j=1}^{N} D_j L^j x_t + G x_t + u_t,$$

where $D_j$ is the $M \times M$ matrix of coefficients on variables lagged $j$ periods, and $G$ is the lower triangular matrix of coefficients on included current variables.[13] In terms of the simple model of equations 4a and 4b, if the money equation were ordered first the two equations would be estimated as

$$(12a) \qquad\qquad m_t = d_{mm}m_{t-1} + d_{mp}p_{t-1} + u_{mt}$$

$$(12b) \qquad\qquad p_t = d_{pm}m_{t-1} + d_{pp}p_{t-1} + g_{pm}m_t + u_{pt}.$$

---

13. This can be seen by substituting each equation of 10 into every equation with a lower order. The $D$ and $G$ matrices are related to $B$ and $C$ matrices by the following identities:

$$G(i, m) = C(i, m) - \sum_{k=1}^{i-1} C(k, m)$$

$$B_j(i, m) = B_j(i, m) - \sum_{k=1}^{i-1} C(k, m)D_j(k, m),$$

where $x(i, m)$ represents the $(i, m)$th element of $x$.

**Table 1. A Three-Equation VAR Model**

| Dependent variable | Number of coefficients on explanatory variables | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| | Lagged | | | Current | | |
| | M | Q | P | M | Q | P |
| M | N | N | N | . . . | . . . | . . . |
| Q | N | N | N | 1 | . . . | . . . |
| P | N | N | N | 1 | 1 | . . . |

Here the money equation contains only lagged values, but the inflation equation also includes the contemporaneous value of money.

The outcome of all this is that when contemporaneous errors have been causally ordered, a VAR model of the form of equation 3 is equivalent to the system of equation 10 or 11, or the simple example of equation 12. And these systems look a lot more like a "conventional" econometric model than 3 because they include both current and lagged values of right-hand variables. The main differences between conventional models and triangularized VAR models are that the latter include all lagged regressors in each equation, impose equal lag lengths, and allow current right-hand variables to enter only in a recursive fashion.

The question remains of how to order the equations. The recursive form (11) suggests that, recalling that $G$ is lower triangular, those variables that respond most to current events, such as changes in exchange rates and interest rates, should be placed at the bottom of the equation list so that their values reflect contemporaneous realizations of variables of a higher order. Conversely, those variables thought by the investigator to be least sensitive to current innovations would be placed at the top; this is consistent with the ordering used by Sims.[14] The ordering chosen clearly depends on the investigator's previous beliefs for, while it seems reasonable to order interest and exchange rates at the bottom of the list, the relative positions of money, output, and prices are controversial.

The implicit appearance of contemporaneous variables in 11 allows us to use a simple tabular device to describe any of the models examined below by indicating which variables contribute coefficients to the $D$ and $G$ matrices. For instance, Sims' simple three-equation model for the

14. Sims, "Macroeconomics and Reality."

levels of money, *M*, output, *Q*, and prices, *P*, can be displayed as in table 1.[15]

The table states that the three equations explaining *M*, *Q*, and *P*, respectively, each contain *N* lagged values of *M*, *Q*, and *P*, while in addition the *Q* equation contains the current value of *M*, and the *P* equation contains the current values of *M* and *Q*. The *M* equation contains no current values. If the right-hand (*G*) matrix were to contain elements above the diagonal, the model would not be recursive and would have to be solved simultaneously.

## Gradual Adjustment of Prices to Demand and Supply Shocks

Whereas the VAR model of the previous section is minimally restricted and atheoretical, this section introduces a more traditional model with many restrictions—both in the construction of variables and in the introduction of particular variables and lag lengths into individual equations—which reflect a mixture of previous beliefs and empirical experimentation. The VAR model is symmetric in variables, whereas the central focus here is on the specification of an equation explaining the rate of change of the aggregate price level. Each additional equation is provided solely to make endogenous a variable that appears on the right-hand side of the inflation equation, rather than for its intrinsic interest. These auxiliary equations are deliberately constructed to avoid the introduction of any additional endogenous variables into the model beyond those appearing in the inflation equation.

### SPECIFICATION OF THE INFLATION EQUATION

The aggregate supply sector of traditional econometric models has typically included two separate equations describing wage and price behavior, with the former including a variable such as the unemployment rate measuring labor market tightness, and the latter involving a variable such as the rate of capacity utilization measuring product market tightness. Yet in the presence of gradual adjustment of wages and prices that is generally assumed in such econometric research, the relevant theoretical framework is a model without market clearing characterized by spillovers between the product and labor market that imply a high correlation between the unemployment of labor and the utilization of

15. Sims, "Interwar and Postwar."

capacity.[16] Indeed, as Okun's law would lead one to expect, the level and change in the ratio of actual to "natural" real GNP (hereafter the output ratio, $\hat{Q}_t$) can explain changes in both wages and prices as well as variables traditionally identified with particular markets, such as the unemployment rate and ratio of unfilled orders to capacity.[17]

The inflation equation developed here is designed to suppress wage changes as both a dependent and independent variable.[18] Wage and price markup equations are specified with restrictions on lags that allow the wage variable to drop out of the model, leaving inflation specified as a function of its own past values, a demand pressure variable, $x_t$, and a vector of various supply shift variables, $z_t$, that may influence the determination of wages, prices, or both:

$$(13) \qquad p_t = \gamma_0 + \gamma_1(L)p_{t-1} + \gamma_2(L)x_t + \gamma_3(L)z_t + \epsilon_t.$$

Here each $L$ in parenthesis indicates that the set of coefficients is allowed to be a polynomial in the lag operator. Each component of the $z$ vector is defined to equal zero when a particular supply shift is absent, allowing a zero value for the sum of the $x_t$ term and the constant term to be interpreted as a "no-shock natural rate" situation compatible with steady inflation ($p_t = p_{t-1}$).

In the research paper that developed the particular form of the inflation equation used here, the proxy for $x_t$ was George Perry's demographically

---

16. The spillover model is analyzed in John Muellbauer and Richard Portes, "Macroeconomic Models with Quantity Rationing," *Economic Journal*, vol. 88 (December 1978), pp. 788–821. The sources of gradual wage and price adjustment are examined in Arthur M. Okun, *Prices and Quantities: A Macroeconomic Analysis* (Brookings Institution, 1981), and Robert J. Gordon, "Output Fluctuations and Gradual Price Adjustment," *Journal of Economic Literature*, vol. 19 (June 1981), pp. 493–530.

17. Robert J. Gordon, "Can the Inflation of the 1970s Be Explained?" *BPEA, 1:1977*, pp. 253–77. The shift from the more structural interpretation of wage and price equations present in Gordon's earlier papers to the present interest in the VAR approach can be traced to those 1977 results and particularly to Christopher Sims's published remarks on that paper (in that same *BPEA* volume, p. 279): "Christopher Sims expressed some amusement that the best wage equation had no labor market variables in it. This result conformed with his belief that wage and price equations cannot be distinguished as applying to different categories of behavior. It was preferable to consider them as interesting statistical reduced-form summaries of the dynamic relationships among the variables."

18. Robert J. Gordon, "Inflation, Flexible Exchange Rates, and the Natural Rate of Unemployment," in Martin Neil Baily, ed., *Workers, Jobs, and Inflation* (Brookings Institution, 1982), pp. 88–155. That paper tests and rejects the inclusion of lagged wages in the wage equation.

weighted unemployment rate, $U_t^W$.[19] The natural weighted unemployment rate can be calculated from 13 as

$$U^{W*} = -\gamma_0 \bigg/ \sum_{j=1}^{N} \gamma_{2j},$$

where the $\gamma_{2j}$ are the individual coefficients in the $\gamma_2(L)$ distribution.[20] In this paper we simplify the presentation by omitting the unemployment rate and substituting the highly correlated log output ratio, $\hat{Q}_t$. Because the natural unemployment rate and the natural real GNP levels are defined by the same criterion, the log output ratio is zero in equilibrium, allowing the constant term to be excluded from 13.[21]

Table 2 presents estimates of 13 for the sample period 1954:2 through 1980:4 and for the first and last halves of the period separately. The estimation for the full sample period allows one parameter change in the middle of the period, a shift in the coefficients on the lagged dependent variable; this sum of coefficients increases modestly in the last half, and the mean lag of the distribution shortens substantially from 14.6 to 8.9 quarters. The shift is highly significant, with $F(4,72) = 4.20$ exceeding the 1 percent critical value of 3.59, and may be due to the increased proportion of workers covered by cost-of-living agreements in the last half of the sample period.

The output ratio entry shows a highly significant sum of coefficients. The remainder of the table lists the sums of coefficients on the various supply-shift variables, $z$. The results for the full sample period in the first column can be summarized as follows. The Nixon-era price controls are estimated to have held down the price level by 1.5 percentage points, and their removal to have raised the price level by 2.5 points. It appears that this estimated effect from removing controls combines the effect of

19. George L. Perry, "Changing Labor Markets and Inflation," *BPEA, 3:1970,* pp. 411–41.

20. Gordon in "Inflation, Flexible Exchange Rates" tests and rejects the hypothesis that the natural weighted unemployment rate shifted upward in the 1970s.

21. Natural real GNP, $Q_t^*$, is set equal to actual real GNP, $Q_t$, in years when the actual weighted unemployment rate was equal to the estimated natural weighted unemployment rate; it is interpolated for intervening years, and is assumed to grow after 1979:1 at an annual rate of 2.75 percent. Our resulting $Q_t^*$ series is $1,520 billion in 1980 and thus is even more pessimistic than the recent $1,546 billion estimate in John A. Tatom, "Potential Output and the Recent Productivity Decline," *Review of the Federal Reserve Bank of St. Louis,* vol. 64 (January 1982), p. 16.

**Table 2. Basic Equation Explaining Quarterly Change in the Fixed-Weight GNP Deflator, Alternative Sample Periods, 1954:2 through 1980:4[a]**

| Independent variable or summary statistic[b] | 1954:2– 1980:4 | 1954:2– 1966:4 | 1967:1– 1980:4 |
|---|---|---|---|
| **Independent variable** | | | |
| Lagged dependent variable, $p_{t-1}$ | | | |
|    1954:2–1966:4 | 0.88** | 0.89* | . . . |
|    Mean lag | (14.6) | (13.7) | |
|    1967:1–1980:4 | 1.01* | . . . | 1.04* |
|    Mean lag | (8.9) | | (7.6) |
| Output ratio, $\hat{Q}_t$ | 0.35* | 0.42* | 0.32* |
| Nixon control dummies, $z_{1t}$ | | | |
|    Controls "on" | −1.49* | . . . | −0.96*** |
|    Controls "off" | 2.47* | . . . | 1.77*** |
| Deviation in productivity growth, $z_{2t}$ | −0.19* | −0.08 | −0.31* |
| Relative price of food and energy, $z_{3t}$ | 0.60* | 0.56 | 0.37 |
| Relative price of imports, $z_{4t}$ | 0.06** | −0.10 | 0.08*** |
| Effective exchange rate for 1975–80, $z_{5t}$ | −0.10* | . . . | −0.07*** |
| Effective minimum wage rate, $z_{6t}$ | 0.03* | 0.04* | 0.04 |
| Effective social security tax rate, $z_{7t}$ | 0.33*** | 0.05 | −0.11 |
| **Summary statistic** | | | |
| $\bar{R}^2$ | 0.956 | 0.859 | 0.940 |
| Standard error of estimate | 0.740 | 0.623 | 0.868 |
| Sum of squared residuals | 39.4 | 8.9 | 18.8 |

Sources: All data are from the national income and product accounts except the effective exchange rate and wage and hourly earnings data, which are from International Monetary Fund, *International Financial Statistics* and U.S. Bureau of Labor Statistics, respectively.

\*    Significant at the 1 percent level.

\*\*   Significant at the 5 percent level.

\*\*\*  Significant at the 10 percent level.

a. The output ratio, $\hat{Q}$, is the log of the ratio of real GNP to natural real GNP. The latter is set equal to real GNP in years when the actual weighted unemployment rate was equal to the estimated natural weighted unemployment rate, is interpolated for intervening years, and is assumed to grow after 1979:1 at an annual rate of 2.75 percent.

The $z_{1t}$ Nixon control dummies are defined to sum to 4.0, since the dependent variable is the quarterly change multiplied by 4.0. Specifically, the Nixon "on" variable is defined as 0.8 for the five quarters 1971:3–1972:3, while the Nixon "off" variable is defined as 0.4 for 1974:2 and 1975:1, and 1.6 for 1974:3 and 1974:4.

The remaining variables are defined as follows: $z_{2t}$—the difference between the rate of growth of nonfarm business productivity and a trend that is allowed to decelerate from 2.56 percent a year during 1956–64, to 2.11 percent for 1964–72, to 1.22 percent for 1972–78, and to 0.5 percent for 1978–81; $z_{3t}$—the rate of growth of the fixed weight personal consumption expenditure deflator minus the growth in the same fixed weight consumption deflator stripped of food and energy; $z_{4t}$—the difference between the rates of growth of the fixed weight import deflator and the fixed weight GNP deflator; $z_{5t}$—the change in the index combining the exchange rates between U.S. dollars and seventeen other major currencies with weights derived from the International Monetary Fund's Multilateral Exchange Rate Model; $z_{6t}$—the difference between the rate of growth of the statutory minimum wage and average hourly earnings in the nonfarm economy; and $z_{7t}$—the percentage change in $(1/(1 - t))$, where $t$ is the ratio of total federal and state and local social security contributions to total wage and salary income. All variables, except for the output ratio and the Nixon control variables, are expressed as rates of change. Quarterly changes are at annual rates.

b. The lagged dependent variable, $p_{t-1}$, is the sum of coefficients of a twenty-four quarter lag distribution constrained to lie along a fourth-degree polynomial with a zero end-point constraint (with mean lags in parentheses); $\hat{Q}_t$ and $z_{3t}$ are the sums of coefficients of an unconstrained lag distribution including the current and four lagged values; $z_{2t}$ is the sum of coefficients of an unconstrained lag distribution including the current and one lagged value; $z_{4t}$, $z_{6t}$, and $z_{7t}$ are the sums of coefficients of an unconstrained lag distribution including four lagged values; and $z_{5t}$ is the coefficient on one lagged value.

ending controls with the cumulative impact of the 1971–74 depreciation
of the dollar, the main effect of which was delayed by the controls until
1974.[22] The coefficient on the deviation of actual productivity growth
from its trend implies that firms base 20 percent of their price-setting
decisions on actual productivity changes, and the remaining 80 percent
on trend productivity growth.[23] Changes in the relative prices of food
and energy are defined as the difference between the growth rates of the
deflator for personal consumption expenditures, respectively including
and excluding expenditures on food and energy. If the dependent variable
were the change in the total consumption deflator, and if the other
explanatory variables influenced only the consumption deflator net of
food and energy with no impact on the difference between the two
deflators, the coefficient on this variable in table 2 would be 1.0. The
actual coefficient of 0.6 results from some combination of, first, the effect
of our choice of the fixed-weight GNP deflator as dependent variable,
particularly the exclusion from this variable of oil and other imports;
and second, the possible negative correlation between other explanatory
variables in table 2, such as the output ratio, and the difference between
the deflators with and without food and energy.

Two other variables, changes in the relative price of imports and in
the effective exchange rate of the dollar, reflect the sensitivity of U.S.
inflation to international events.[24] Last, the equation includes two
domestic supply-shift variables, changes in the effective minimum wage
rate and in the effective social security tax rate. The coefficient on the
latter indicates that about one-third of an increase in the combined
payroll tax (employee plus employer share) is shifted forward to prices,
and the burden of the remainder falls on profits and wages.

22. This interpretation is explained in Gordon's "Inflation, Flexible Exchange Rates"
as due to the fact that the exchange rate is allowed to have an impact only beginning in
1975:2. Thus the controls "off" coefficient combines the effect of ending controls with the
cumulative impact of the 1971–74 depreciation of the dollar.

23. The productivity growth trend is allowed to decelerate from 2.56 percent a year
during 1954–64 to 2.11 percent for 1964–72, to 1.22 percent for 1972–78, and to 0.5 percent
for 1978–81. The estimated coefficient on the productivity growth deviation of $-0.19$ is
remarkably close to the figure of $-0.24$ estimated more than a decade ago in Robert J.
Gordon, "Inflation in Recession and Recovery," *BPEA, 1:1971,* p. 129.

24. The former is defined as the difference between the quarterly rates of change of
the fixed-weight import deflator and fixed-weight GNP deflator. The latter is defined as the
change in the effective exchange rate using the IMF Multilateral Exchange Rate Model
weights; see International Monetary Fund, *International Financial Statistics,* line am.x.

STRUCTURE OF THE MODEL

The small econometric model designed to calculate the output and price effects of a monetary deceleration adds to 13 the minimum number of equations needed to explain its endogenous explanatory variables. Unlike the VAR approach, in which all variables are usually treated as endogenous, here some of the relevant variables are assumed to be exogenous:

| *Endogenous* | *Exogenous* |
| --- | --- |
| Food-energy effect, $z_{3t}$ | Adjusted money-supply growth, $\hat{m}_t$ |
| Change in relative price of imports, $z_{4t}$ | Price control dummies, $z_{1t}$ |
| | Change in effective minimum wage, $z_{6t}$ |
| Adjusted nominal GNP growth, $\hat{y}_t$ | |
| Output ratio, $\hat{Q}_t$ | Change in effective social security payroll tax, $z_{7t}$ |
| Deviation in productivity growth, $z_{2t}$ | |
| Inflation rate, $p_t$ | |
| Change in U.S. effective exchange rate, $z_{5t}$ | |

The endogenous variables are arranged in an order that treats the food-energy effect and relative price of imports as "most exogenous" and allows the inflation rate and effective exchange rate to be influenced by current innovations in each of the variables listed above them. The variables included in each equation are shown in table 3, which has a format similar to that of table 1.

The first two variables listed, the food-energy effect and the relative price of imports, are often treated as exogenous. Here each of the two is allowed to depend on its own lagged values, the lagged values of the other, and the lagged effective exchange rate. Money, nominal GNP, and inflation are excluded from the equations for these two variables because in simulations of future policies we do not want the rate of relative price change to be influenced permanently by changes in the growth rates of nominal money and GNP.[25]

25. Although the effective exchange rate is also a nominal variable, the equation describing its determination is neutral in the long run with respect to changes in the growth rate of nominal money. Our justification for this specification is given below.

**Table 3. The Basic Restricted Model[a]**

| Dependent variable | Lagged coefficients — Endogenous $z_3$ | $z_4$ | $\hat{y}$ | $\hat{Q}$ | $z_2$ | $p$ | $z_5$ | Exogenous $\hat{m}$ | $z_6$ | $z_7$ | Current coefficients — Endogenous $z_3$ | $z_4$ | $\hat{y}$ | $\hat{Q}$ | $z_2$ | $p$ | $z_5$ | Exogenous $\hat{m}$ | $z_1$ | $z_8$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Relative price of food and energy, $z_3$ | ... | 4 | ... | ... | ... | ... | 4 | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| Relative price of imports, $z_4$ | 4 | 4 | ... | ... | ... | ... | 4 | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| Adjusted nominal GNP growth, $\hat{y}$ | ... | ... | 4 | ... | ... | ... | ... | 4 | ... | ... | 1 | ... | ... | 1 | ... | ... | ... | 1 | ... | ... |
| Output ratio, $\hat{Q}$ | ... | ... | ... | 1 | ... | ... | ... | ... | ... | ... | ... | ... | 1 | ... | ... | 1 | ... | ... | ... | ... |
| Productivity growth deviation, $z_2$ | 2 | ... | ... | 4 | ... | ... | ... | ... | ... | ... | 1 | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| Fixed-weight GNP deflator, $p$ | 4 | 4 | ... | 4 | 1 | 24 | 1 | 8[b] | 4 | 4 | 1 | 1 | 1[b] | ... | ... | ... | ... | 1[b] | 1 | ... |
| Effective exchange rate, $z_5$ | ... | ... | 4[b] | ... | ... | ... | 4 | ... | ... | 4 | 1 | 1 | ... | ... | ... | ... | ... | ... | ... | 1 |

a. See definition of variables in table 2, note a. The growth rates of M1 and nominal GNP, less the growth rate of natural real GNP, are denoted by $\hat{m}$ and $\hat{y}$, respectively. The dummy variables in the exchange rate equation are indicated by $z_8$ and defined as $D72 = 1.0$ in 1972:1 and zero otherwise and as $D73 = 1.0$ in 1973:1 and 1973:2 and zero otherwise.

b. For this equation $\hat{m}$ is actually $\tilde{m}$, the deviation of actual money growth from its three-year moving average, and $\tilde{y}$ is the deviation of velocity growth from its long-run trend. See the discussion of equation 15 in the text.

The sums of coefficients for these equations and those for the nominal GNP and labor productivity equations are set out in table 4. It can be seen that the relative price of food and energy, $z_3$, depends most significantly on the foreign exchange rate, $z_5$. By contrast the relative price of imports, $z_4$, depends little on the exchange rate directly, but is very strongly influenced by its own lagged value and by the food-energy variable. The high coefficient on current and lagged food and energy prices appears to be due to the unusual and correlated movements of oil prices, import prices, and the exchange rate in the 1970s. In view of the possible spuriousness of this coefficient for long-run simulations, we later examine the sensitivity of the results of our model's simulation to the exclusion of the food-energy and import price equations.

Because the model is designed to trace the output and price effects of alternative deterministic monetary growth paths, money growth is treated as an exogenous variable. The growth rates of money and the nominal GNP are adjusted by netting out the growth of natural real GNP ($\hat{m}_t = m_t - q_t^*$; $\hat{y}_t = y_t - q_t^*$). This allows us to move back and forth between these nominal growth rates and the output ratio, using the basic identity,

$$(14) \qquad \hat{Q}_t \equiv \hat{Q}_{t-1} + \hat{y}_t - p_t.$$

To avoid introducing any additional variables relevant to the determination of aggregate demand, the adjusted growth rate of nominal GNP is determined in a bivariate Granger-type VAR equation in which the only explanatory variables are lagged values of adjusted nominal GNP and current and lagged adjusted money growth. Then the inflation equation 13 and the identity 14 are solved simultaneously to split current nominal GNP growth between inflation and changes in the output ratio.

This leaves three endogenous variables in table 3 to be determined. Deviations in productivity growth from trend (row 5) depend on lags in firing and hiring, which make the productivity variable a function of current and past changes in the output ratio.[26] The productivity variable is also allowed to be influenced by the food-energy effect. The coefficient sums in table 4 show the food-energy effect on productivity, but mask

---

26. This specification and the timing of the slowdown in the trend are consistent with the empirical description of cyclical productivity effects in Robert J. Gordon, "The 'End-of-Expansion' Phenomenon in Short-Run Productivity Behavior," *BPEA*, 2:1979, pp. 447–61.

Table 4. Auxiliary Equations for the Basic Model, 1954:2 through 1980:4[a]

| Sums of coefficients on current and lagged variables and summary statistic | Dependent variable | | | |
|---|---|---|---|---|
| | Relative price of food and energy, $z_3$ | Relative price of imports, $z_4$ | Adjusted nominal GNP growth, $\hat{y}$ | Productivity growth deviation, $z_2$ |
| Sums of coefficients | | | | |
| $z_3$ | 0.10 | 3.36* | . . . | −0.43 |
| $z_4$ | 0.05 | 0.35* | . . . | . . . |
| $\hat{y}$ | . . . | . . . | −0.14 | . . . |
| $\hat{Q}$ | . . . | . . . | . . . | −0.10* |
| $z_5$ | −0.17* | −0.02 | . . . | . . . |
| $\hat{m}$ | . . . | . . . | 1.27* | . . . |
| Summary statistic | | | | |
| $\bar{R}^2$ | 0.53 | 0.61 | 0.52 | 0.62 |
| Standard error of estimate | 0.95 | 5.1 | 3.1 | 2.0 |

Source: Same as table 2.
*Significant at the 1 percent level.
a. See the definitions of variables in table 2, note a, and table 3, note a.

the influence of output movements that primarily influence productivity in proportion to the rate of change of the output ratio rather than to its level. The actual coefficients imply that a 1 percent increase in the output ratio would be associated with a transitory 2.4 percent increase in productivity, which is then reversed in the following five quarters.

The inflation equation (row 6) is the same as that displayed in the first column of table 2. The specification of changes in the foreign exchange rate is quite unconventional, as it is motivated by a desire to keep interest rates and foreign money and income variables out of the model. Clearly, the exchange rate should appreciate in response to a deceleration in domestic money growth, but a constraint is needed in future simulations to keep the exchange rate from appreciating forever.

The equation summarized in row 7 of table 3 introduces $\tilde{m}_t$, the deviation of actual money growth from its three-year moving average, where the latter may be considered a proxy for foreign money growth and represents the idea that a monetary deceleration in the United States will be followed in due course by a deceleration in foreign money growth. The deviation of velocity growth from its long-run trend ($y_t - m_t - 3.2$) is also included, along with the food-energy variable and two

dummy variables for the sharp correction in the overvaluation of the dollar that occurred after the Smithsonian Agreement and in early 1973. The estimated equation for the 1972:1–1980:4 sample period is

$$(15) \quad z_{5t} = \sum_{i=1}^{9} \beta_i \bar{m}_{t-i+1} + \sum_{i=1}^{5} \delta_i (y_{t-i+1} - m_{t-i+1} - 3.2)$$

$$+ \ 2.07^{**} z_{3t} - 16.6^{*} D72 - 33.6^{*} D73$$

$$R^2 = 0.75, \text{ standard error} = 7.5, \Sigma\beta_i = -4.3^{**}, \Sigma\delta_i = 1.4^{**}.$$

where $\bar{m}_t = m_t - (1/12)\sum_{i-1}^{12} m_{t-i}$, $D72 = 1.0$ in the first quarter of 1972 (0 otherwise), and $D73 = 1.0$ in the first and second quarters of 1973 (0 otherwise), and the asterisks have the same meaning as in table 2. The resulting equation for exchange rate has the property that the exchange rate appreciates while money growth decelerates, but reaches a new steady-state level when money growth arrives at its final constant growth rate in the simulations reported below.

　　Overall, the model is similar in structure to those that have been used to simulate the effects of monetary policy in our previous work. The main innovation here is the treatment of the food-energy and import variables as endogenous.[27] Compared to an unconstrained VAR model including the same variables, the main justification for the many zero restrictions in the model shown in table 3 is a conscious attempt to separate real from nominal effects, so that the numerous variables representing relative price changes approach zero in the long run in simulations of alternative nominal money-growth paths.

### Conceptual Issues in Simulating Future Policies

THE LUCAS CRITIQUE

　　Some economists, following the lead of Robert Lucas, object to econometric simulations of hypothetical future policy actions based on

---

　　27. A similar model is displayed in Gordon, ''Inflation in Recession and Recovery,'' appendix B. A similar treatment of the foreign exchange rate was introduced in ''Inflation, Flexible Exchange Rates,'' equation A.5.

parameters estimated from a historical sample period when a different policy regime may have been in effect. In the specific case we examine, the dynamic response of output and price adjustment to a hypothetical future monetary deceleration depends mainly on the parameters in an inflation equation estimated for the 1954–80 sample period. Critics might argue that the inflation equation 13 is misspecified; in place of the $\gamma_1(L)$ lag distribution on past inflation should be substituted the expected rate of inflation, say $Ep_t$. Because no sustained monetary deceleration was ever actually carried out within the sample period, they would claim we have no evidence to rule out a much more prompt response of $Ep_t$ to the announcement of a monetary deceleration in 1981—the Volcker policy—than would be indicated by the historical lag distribution on past inflation rates.

Our willingness to take seriously simulations of hypothetical future monetary policies rests on the parallel nature of the hypothetical 1981–86 monetary deceleration and the actual 1965–70 monetary acceleration. Our argument in the following paragraphs can be divided into three components. There was a monetary ''regime shift'' in the mid-1960s that was more significant statistically than that implied by the Volcker monetary slowdown. Economic agents would have taken several years to recognize a regime shift in the mid-1960s and, presumably, a shift in the opposite direction in 1981. And, perhaps most important, the structure of our basic inflation equation exhibits structural stability when estimated across two subperiods (1954–66 and 1967–80) that bracket the mid-1960s monetary regime shift, thus yielding no presumption that a structural shift in that equation would occur in the early 1980s, even after the several years that would elapse before such a shift could be recognized.

In the literature on the Lucas critique a monetary regime is in effect over a given time interval if the evolution of monetary growth can be described by a feedback rule having stable parameters. It is taken for granted in existing studies of monetary regime shifts, such as that of Thomas Sargent and Salih Neftci, that a change in regime can be identified econometrically by applying a Chow test to an equation in which the growth rate of the money supply is the dependent variable, and both the lagged dependent variable and other key macroeconomic aggregates to which the monetary authority might react are on the right-hand side of

the equation.[28] Indeed, when equations explaining quarterly M1 growth for the three alternative sample periods 1954–80, 1954–66, and 1967–80 are estimated with four lags of M1 growth, inflation, and the output ratio as explanatory variables, a Chow test confirms that a structural shift at the beginning of 1967 is significant; the $F(13,81)$ ratio is 2.26, compared to a 5 percent critical value of 1.87.

Is such a shift in structure implied by the deterministic money growth paths used to generate the post-1980 simulations in the next section? When the 1981–92 series of assumed money growth paths are treated as a dependent variable, and the generated values of inflation and the output ratio are treated as explanatory variables (along with the lagged dependent variable), a Chow test comparing the stability of 1981–92 coefficients with 1967–92 coefficients reveals no shift in structure in either the control or Volcker solutions. For the control solution path the $F(13,91)$ ratio is 0.38, compared to a 5 percent critical value of 1.86, and for the Volcker solution the analogous $F$-ratio is 1.73.

The Lucas critique implies that a recognized shift from the stable parameters in one monetary regime to another set of stable parameters for a second monetary regime should lead to an instantaneous shift in the behavior of private agents. Yet a crucial flaw in this argument is the assumption of instant recognition that a regime change has occurred: how does one recognize such a change?

Consider the monetary regime shift at the beginning of 1967, described in the previous section, that can be recognized by the econometrician performing tests on data available in 1982. Could such a shift have been recognized and thus have been a source for a behavioral parameter change in 1967? Using currently available data, we can compare M1 equations estimated for the full period from 1954 to the end of the year $L$ and two equations extending from 1954 to $L - 5$ and from $L - 5$ to $L$. We find that, while the Chow test reveals a structural shift significant at the 10 percent level as early as $L = 1968$, a structural shift is identified using the more conventional 5 percent significance level only when several more years have passed and $L = 1972$. With such flimsy evidence

28. Salih Neftci and Thomas J. Sargent, "A Little Bit of Evidence on the Natural Rate Hypothesis from the U.S.," *Journal of Monetary Economics*, vol. 4 (April 1978), pp. 315–19.

available in the interim between 1968 and 1972, our hypothetical "yeo-man agent-econometrician" would have had no firm reason for changing price- and wage-setting practices and institutions in the interim.[29]

These yeoman agent-econometricians would not only have trouble distinguishing a regime change if one were to occur in 1981, but would also have no precedent for shifting their wage-setting and price-setting behavior in response to such a shift. Table 2 presents inflation equations for the 1954–66 and 1967–80 subperiods corresponding to the apparent monetary regime change that occurred at the beginning of 1967.[30]

What seems remarkable to us is, despite a few minor exceptions, the overall stability of the sums of coefficients in the inflation equation across the two subintervals. The only significant coefficient shift, that in the distribution on the lagged dependent variable, is already included in the full-period equation. A Chow test confirms that there is no significant change in structure when the other coefficients are allowed to shift in 1967:1; the $F(24,48)$ ratio is 0.85, compared to a 10 percent critical value of 1.53.

Thus, even if the post-1980 monetary deceleration were sufficiently dramatic to be interpreted by agents, perhaps with a lag of two to five years, as a regime change, we are left with no solid reason to think there would be a marked change in the structure of the inflation equation, and thus in the estimated "sacrifice ratio." The only change in the inflation process after 1967 was a shortening of the lag distribution. If a policy

29. The phrase "yeoman agent-econometrician" combines three essential elements of the new classical equilibrium macroeconomics associated with the names of Lucas, Sargent, and Barro. First, their microeconomic behavioral models are most appropriate for price-taking "yeoman farmers," as pointed out by Alan S. Blinder and Stanley Fischer in "Inventories, Rational Expectations and the Business Cycle," *Journal of Monetary Economics,* vol. 8 (November 1981), pp. 277–304. Second, the individuals making decisions in the Lucas, Sargent, Barro literature are almost always described as agents. Third, the reliance of such yeoman agents on Chow tests to identify shifts in regimes implies that they have all received a rudimentary education in econometrics.

30. We identify such a policy shift by a significant change in parameters, not by an explicit announcement of a policy shift by the Board of Governors. The 1967 shift involved a significant reduction of the previously negative coefficient on the inflation rate in the M1 equations; this suggests that the Federal Reserve's behavior shifted through its failure significantly to decelerate monetary growth in response to the upsurge of inflation that occurred in the 1966–68 interval. It is unlikely that a new policy involving the "failure to fight inflation" would have been announced explicitly by the Board.

shift caused history to "rewind" to the longer lag distribution in effect before 1967, our simulations would be too optimistic, not too pessimistic as the critics suggest.[31]

Some skeptics may resist the preceding analysis, which follows Neftci-Sargent by basing its assessment of regime shifts entirely on the behavior of the money supply. Instead one could examine the behavior of interest rates and might conclude that the willingness of the Federal Reserve to tolerate high interest rates since its announced November 1979 policy shift, despite relatively high unemployment, is unprecedented.[32] The widespread wage concessions and contract renegotiations of 1981–82 seem consistent with widespread perception of a new toughness in the Federal Reserve's stance. Yet the implication that our simulations may be too pessimistic is not supported by the data for 1981:1 through 1982:1. Our basic model of table 3, row 1, when simulated using the actual monetary growth rates between 1981:1 and 1982:1, underpredicts both the inflation rate (an average predicted rate of 7.7 percent versus the actual 8.0 percent) and the unemployment rate (predicted 7.4 percent versus the actual 7.8 percent).[33]

31. Cross-country historical evidence from the last century suggests that inflation inertia is a unique phenomenon of the postwar United States and that the timing of parameter shifts is consistent with the view that the institution of staggered three-year wage contracts is the main culprit. There is no reason to believe that a drastic shift would occur in the structure of the inflation equation until a regime shift far more drastic than that in 1967 were to cause multiyear contracts to be abandoned. See Robert J. Gordon, "Why U.S. Wage and Employment Behavior Differs from that in Britain and Japan," *Economic Journal,* vol. 92 (March 1982), pp. 13–44.

32. As evidence that the shift in Federal Reserve interest rate policy in November 1979 was not perceived to be permanent, we can cite Fair's published belief that the policy shift was temporary and had ended in mid-1980. See Ray C. Fair, "Estimated Effects of the October 1979 Change in Monetary Policy on the 1980 Economy," *American Economic Review,* vol. 71 (May 1981 *Papers and Proceedings, 1980*), pp. 160–65.

33. The error in predicting unemployment was quite large in 1982:1, when the basic model predicted an unemployment rate of 7.6 percent along the Volcker path in figure 1, as contrasted with the 8.8 percent rate that actually occurred in that quarter. Most of this error is caused by our simplistic equation that translates money growth into nominal GNP growth, not by the inflation equation itself that predicts inflation given unemployment. The forecast error of 1.2 percentage point of unemployment can be decomposed as follows: actual quarterly path of M1 growth in 1981 in contrast to the constant 5.0 percent rate assumed in figure 1, 0.2 extra point of unemployment; slowdown in velocity growth not predicted by nominal GNP equation, 0.6 point; underprediction of unemployment rate, with corresponding overprediction of output ratio, 0.3 point; and error in unemployment equation for a given output ratio, 0.1 point.

Finally, we can concede that the structure of the inflation process might change in some unpredictable way after sufficient time, say five years, has passed for a monetary regime shift to be identified; indeed, the twelve year interval between 1981 and 1992 *is* a long time to look into the future. But a structural change after five years would not alter our conclusion that stopping inflation is costly simply because most of the output cost occurs early in the simulation interval (91 percent of the cost occurs in the first five years along path I in figure 1 below, and 56 percent along path II).

### THE SACRIFICE RATIO

Arthur Okun computed the output loss from reducing inflation implied by a number of Phillips curve models and came up with estimates of the output cost of reducing inflation by one percentage point of between 6 and 18 percent of a year's GNP, with a mean of 10 percent.[34] Those estimates were based on a ratio between the loss of output, in percent of GNP, and the reduction in inflation, in percentage points, occurring in the first year of a disinflation experiment. This method of calculation does not, however, take into account the possibility of changes in the ratio as the disinflation experiment proceeds.

Here we investigate a disinflationary monetary strategy and calculate the ratio of the present discounted value of the cumulative output loss to the average discounted reduction in inflation. While these might, in principle, be computed for an infinite time horizon, we economize on computation cost by calculating the ratio of terms discounted forty-eight quarters into the future as

$$(16) \qquad S^{10} = \frac{\sum_{t=1}^{48} (\hat{Q}_t^1 - \hat{Q}_t^0)/4(1 + r)^t}{\left( \sum_{t=1}^{48} (p_t^1 - p_t^0)/(1 + r)^t \right) \Big/ \sum_{t=1}^{48} (1 + r)^{-t}},$$

where the superscript 1 refers to a control simulation and 0 refers to a simulation perturbed by a deterministic money-growth deceleration. In 16 the cumulative output loss is divided by four to convert it to an annual

34. Okun, "Efficient Disinflationary Policies."

**Figure 1. Effects of Different Monetary Policies on Unemployment, Inflation and the Sacrifice Ratio, Alternative Models, 1981–1992[a]**



Percent

Unemployment

Percent

Inflation

Percentage points or ratio

Sacrifice ratio and inflation displacement[b]
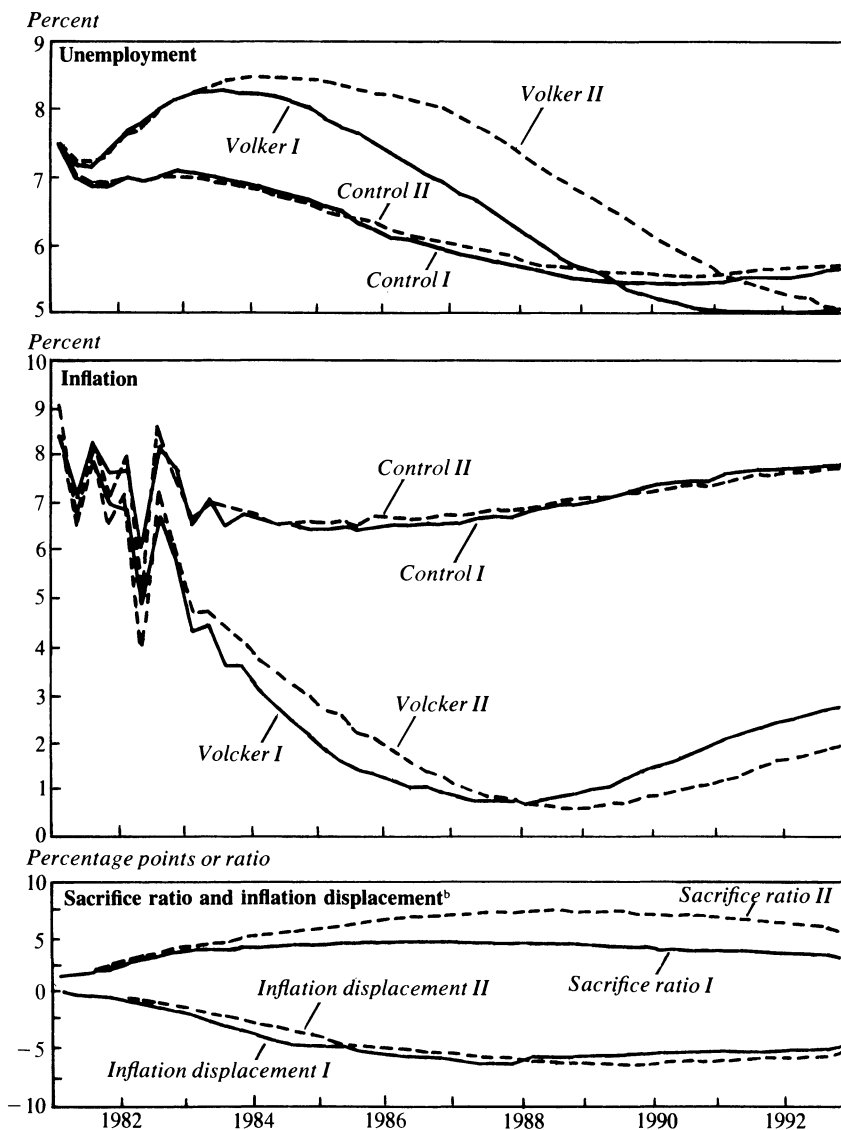
a. Path I is the one presented as the basic model in table 5, row 1. Path II holds constant the relative prices of imports and food and energy, as presented in table 5, row 3.

The control solution sets the growth rate of M1 at 6.6 percent a year. The Volcker solution sets 5.0 percent for 1981, 4.0 percent for 1982, and then decelerates by 0.5 point a year to a rate of 2.0 percent for 1986 through 1992.

b. The sacrifice ratio is undiscounted. The inflation displacement is the difference in the inflation path between the control and Volcker solutions.

basis, and the denominator is divided by $\sum_{t=1}^{48} (1 + r)^{-t}$ in order to average the inflation rate, so that, for example, if $p_t^1 - p_t^0$ were constant at a rate $\pi$, the denominator would just equal $\pi$.

Obviously the choice of discount rates is crucial once we use a procedure that takes account of developments over several years. The analogous procedure to Okun's would be to ignore the relative timing of costs and benefits and simply to divide the cumulative output loss after twelve years by the permanent reduction in inflation. We report results on this basis ($r = 0$) and also with a positive annual discount rate ($r = 3$), which provides a better starting point for welfare analysis.

An important issue raised by this set of calculations involves the limitation of the horizon to twelve years. As we show below, our disinflationary monetary strategy overshoots the equilibrium output ratio and inflation rate by varying amounts in the different simulations, and in most cases the economy has not settled down by the end of 1992. This causes an overstatement in our sacrifice ratio by excluding the discounted benefit of lower inflation after 1992, as well as any possible increase in the growth rate of "natural" output, which is assumed below to be exogenous. It also, however, understates the sacrifice ratio by failing to include the post-1992 recession that arises from overshooting, the cost to society of the instability in both output and inflation that is caused by the disinflationary strategy, and any diminution in the capital stock due to low investment during the 1981–85 slump. We assume that the net effect of these distortions is small enough so that our results are not significantly biased. As further justification for a truncated horizon, we feel that it is unwise to give too much weight to the parts of the simulation that are remote in time from the starting date and therefore subject to large forecasting errors.[35]

35. For a more detailed discussion of the welfare costs of disinflation, see Okun, *Prices and Quantities,* chaps. 7 and 8. In particular, we follow Okun in treating a positive log output ratio as creating a benefit for society, due to the role of the tax "wedge" that makes labor's marginal product exceed its opportunity cost at a zero log output ratio. For a detailed analysis, see Robert J. Gordon, "The Welfare Cost of Higher Unemployment," *BPEA, 1:1973,* pp. 133–95. A comprehensive analysis of the costs of inflation is contained in Stanley Fischer, "Towards an Understanding of the Costs of Inflation: II," in Karl Brunner and Allan H. Meltzer, eds., *The Costs and Consequences of Inflation,* Carnegie-Rochester Conference Series on Public Policy, vol. 15 (Amsterdam: North-Holland, 1981), pp. 5–41.

It is instructive to consider the implications of a discounted sacrifice ratio of, say, 6. Such a ratio would imply that in order to achieve a long-run reduction in the inflation rate of 5 percentage points, the economy would have to sacrifice output with a present value of 30 percent of a year's natural GNP, roughly $1,000 billion at current prices, or about $4,000 per capita.

However large, the output loss from disinflation does not by itself contain implications for economic policy. An assessment must be made of the welfare cost of lost output and the welfare benefit of lower inflation. Consideration of the value of the leisure time gained by the unemployed reduces the loss of $1,000 billion in domestic output to about $860 billion.[36] Lowering inflation would yield benefits to society reflecting the nonneutral impact of financial regulation and the tax system. For example, Stanley Fischer estimates the annual gain from a 5 percentage point reduction of inflation as 0.30 percent of GNP. This reflects reduced distortion in holdings of noninterest-bearing money and interest-bearing assets subject to interest rate ceilings.[37] The gain from lower inflation can be boosted to as much as 1 percent of GNP by considering the effects of inflation on saving, although all of this added effect hinges on the assumption that tax reform is infeasible. Were the total annual gain from reducing inflation by 5 points to amount to as much as 1 percent of GNP ($30 billion), the present value of the gain from reducing inflation would be $1,000 billion, exceeding the present value of the output loss of $860 billion. But we do not believe that tax distortions should be treated as unalterable and permanent.

THE CONTROL AND VOLCKER SOLUTIONS

To carry out a simulation whose results are directly relevant to the contemporary policy debate, we compare a control solution with an approximation of the current official policy of the Federal Reserve Board. The control solution sets the annual growth rate of M1 permanently at its 1980 average of 6.6 percent a year. The alternative disinflationary Volcker solution sets 1981 growth at the actual average of 5.0 percent, sets 1982 growth at the midpoint of the official target range, 4.0

36. Gordon, "The Welfare Cost of Higher Unemployment," p. 164.
37. Fischer, "Towards an Understanding," pp. 17–19.

percent, and then allows M1 growth to decelerate by 0.5 percentage point a year to a final rate of 2.0 percent for 1986 through 1992.[38]

An alternative to this comparison of solutions would be the "innovation accounting" approach generally used in the evaluation of VAR models. A downward innovation in M1 growth could be introduced in the first quarter of the simulation, equal in size to one sample-period standard deviation, and the subsequent adjustment of the output ratio and inflation rate could be calculated. Because the shock occurs for only one period, the resulting sacrifice ratio would differ from that in the control and Volcker simulations because there would be more time for the overshooting cycles to dampen. These differences are difficult to explain in a compact way, however, and we choose to limit the size and complexity of the paper by presenting simulation results only for the control and Volcker alternatives.

### Sacrifice Ratios in Alternative Models

Table 5, which summarizes the simulation results and implied sacrifice ratios, is divided into two sections. Rows 1 through 4 use alternative versions of the basic model from table 3. The remainder of the table shows how the results are altered when we convert our basic model, step by step, into a simple VAR model. Each line of the table displays goodness-of-fit statistics for the inflation equation in the first two columns, the undiscounted cumulative twelve-year output loss in the third column, and the average reduction of the inflation rate in the fourth column. The last two columns show the sacrifice ratio from equation 16 with the undiscounted ratio ($r = 0$) and the sacrifice ratio discounted at an annual rate of 3 percent ($r = 3$)

The first row of the table shows that the basic model of table 3 generates a cumulative output loss of 13.4 percent of a year's GNP to reduce inflation by an average of 4.4 percentage points a year for sacrifice ratios of 3.0 (undiscounted) and 4.3 (discounted). Discounting raises the ratio, of course, because the output loss comes relatively early in the 1981–92 period, and the benefit of lower inflation comes later. The

---

38. The 1980 and 1981 actual figures are fourth-quarter to fourth-quarter, as reported in Board of Governors of the Federal Reserve System, "Monetary Policy Objectives for 1982," February 10, 1982, pp. 6–7. The 1981 "shift adjustment" of M1B is ignored.

**Table 5. Alternative Models: Sample Period Goodness-of-Fit Statistics for the Inflation Equation and Implied Sacrifice Ratios, 1981–82**

| Model description[a] | Inflation equation in sample period | | Control versus Volcker money path[b] | | | Discounted sacrifice ratio[b] |
|---|---|---|---|---|---|---|
| | Sum of squared residuals | Standard error of estimate | Cumulative output loss (percent) | Average inflation reduction (percentage points) | Sacrifice ratio | |
| **Basic model** | | | | | | |
| 1. Same model as table 3 | 39.4 | 0.74 | 13.4 | 4.4 | 3.0 | 4.3 |
| 2. $z_4$ exogenous | 39.4 | 0.74 | 21.0 | 4.4 | 4.8 | 5.8 |
| 3. $z_3$, $z_4$ exogenous | 39.4 | 0.74 | 27.5 | 4.5 | 6.2 | 7.2 |
| 4. $z_3$, $z_4$, $z_5$ exogenous | 39.4 | 0.74 | 37.9 | 4.5 | 8.4 | 9.9 |
| **Symmetric VAR models[c]** | | | | | | |
| 5. Variables same as in table 3[d] | 46.1 | 0.90 | −0.5 | 4.3 | −0.1 | 1.7 |
| 6. Variables same as preceding row, excluding $\hat{m}$ from inflation equation | 47.3 | 0.90 | −0.5 | 4.3 | −0.1 | 1.8 |
| 7. Variables same as in table 3, four lags on all variables | 71.8 | 0.98 | −0.3 | 4.1 | −0.1 | 1.2 |
| 8. Same as preceding row, with reverse ordering | 82.5 | 1.04 | −7.2 | 4.3 | −1.7 | 0.0 |
| 9. Only $\hat{m}$, $\hat{Q}$, $p$, $z_4$ included | 94.1 | 1.03 | −3.8 | 4.6 | −0.8 | 1.0 |
| 10. Only $m$, $q$, $z_4 + p$ included | 95.6 | 1.04 | 24.5 | 4.5 | 5.5 | 6.1 |
| 11. Only $M$, $Q$, $P$, $PZ_4$ included | 103.4 | 1.08 | 144.7 | 3.5 | 41.7 | 34.2 |

Source: Simulations by the authors.

a. Variables are defined as in table 2, note a, and table 3, note a. In models 10 and 11, $q$ indicates the rate of growth in actual real GNP, and capital letters indicate the levels of the variables.

b. For a definition of the sacrifice ratio, see the text. The discounting was done at an annual rate of 3 percent.

c. In VAR models each variable appears in each equation with four lags except where noted below.

d. Twenty-four quarter lagged dependent variable and lags zero through four on $\hat{m}$ in the inflation equation.

permanent reduction in the inflation rate in equilibrium is 4.9 percent, and it is accompanied by a 4.6 percentage point reduction in M1 growth and a 0.3 point reduction in velocity growth.

The economy's dynamic adjustment is illustrated in figure 1, where the solid lines indicate the simulations being discussed here. Because the unemployment rate is a more familiar statistic than the output ratio, we display in the top panel the implied unemployment rate profile for 1981–92 under the control and Volcker simulations.[39] Whereas the control unemployment rate remains in the range of 5.5 to 7.4 percent, the Volcker unemployment rate peaks at 8.3 percent in 1983 and then drops rapidly to a trough of 5.0 in 1991, substantially overshooting its natural rate of 6.0 percent and implying additional instability for the post-1992 period. In the second panel the relatively stable control inflation rate is contrasted with the plummeting Volcker inflation rate, which hits a trough of 0.7 percent in 1987. The last panel shows the undiscounted sacrifice ratio and the displacement of the inflation rate between the control and Volcker projections.[40] If the variance rather than the mean of the inflation rate is what matters for its welfare cost, a defect of the Volcker policy is the extra instability that it creates for inflation over this period.

INTERNATIONAL EFFECTS

Why is the sacrifice ratio on row 1 of table 5, both with and without discounting, so much lower than the ratio of 10 reported in Okun's survey? Our more optimistic set of results reflects three channels of "international feedback" included in the basic model. The Volcker

---

39. The unemployment rate is calculated from the following Okun's Law equation estimated in Gordon's "Inflation, Flexible Exchange Rates":

$$U_t^w = 3.96 - 0.243\hat{Q}_t - 0.142\hat{Q}_{t-1} - 0.040\hat{Q}_{t-2}$$
$$(46.2) \quad (-12.0) \quad (-6.39) \quad (-1.78)$$

$$\bar{R}^2 = 0.976,\ \text{Durbin-Watson} = 1.55,\ \text{Standard error} = 0.178,$$

where numbers in parentheses are *t*-ratios.

To convert from the weighted to the official unemployment rate, the constant is changed from 3.96 to 6.00 percentage points.

40. The plotted undiscounted sacrifice ratio is based on a separate calculation for each period. Thus the plotted value for 1992:4 corresponds to that listed in the fifth column of table 5.

simulation reduces the inflation rate not only through the traditional channel of lower output, but also by causing a reduction in the relative price of imports and in the relative price of food and energy, as well as an appreciation in the effective exchange rate. The impact of these channels of monetary influence is demonstrated in table 5. In row 2 the relative import price change variable, $z_4$, is set at zero during the 1981–92 simulation, in contrast to its endogenous response allowed in row 1. The consequence of imposing exogeneity on the $z_4$ variable is an increase in the discounted sacrifice ratio from 4.3 to 5.8. In parallel fashion, row 3 treats both the relative import price and food-energy, variables $z_4$ and $z_3$, as exogenous, raising the discounted sacrifice ratio to 7.2. Finally, in row 4 all three international feedback variables are made exogenous in the simulation, resulting in a discounted sacrifice ratio of 9.9 that is close to Okun's summary estimate of 10.

Since the endogeneity of the international variables accounts for the more optimistic results in row 1 as compared to row 4, we may ask whether the behavior of the international variables in the two simulations, as summarized in the following, is plausible:

|  | Cumulative changes, 1981–92 (percent) | | |
|---|---|---|---|
|  | Control | Volcker | Difference |
| Food-energy effect | − 3.8 | − 6.0 | 2.2 |
| Relative price of imports | − 7.7 | − 32.3 | 24.6 |
| Effective exchange rate | 4.6 | 20.8 | 16.2 |

Although the food-energy change seems minor, the exchange rate difference of 16.2 is substantial. It is quite close, however, to the 13.7 percent cumulative appreciation of the same exchange rate measure that actually occurred between 1980:4 and 1981:4. Since the cumulative displacement of the domestic price level between the two simulations is 53.2 percent, the exchange rate results would be consistent with the long-run achievement of purchasing power parity if the Volcker policy caused a cumulative displacement of the foreign price level by 37 percent (53.2 minus 16.2), that is, by about two-thirds of the U.S. displacement. In this case, by 1992 the real U.S. exchange rate would have returned to its 1980 value.[41]

41. See the related discussion in Willem H. Buiter and Marcus Miller, "Real Exchange Rate Overshooting and the Output Cost of Bringing Down Inflation," in *European Economic Review*, vol. 18 (May–June 1982), pp. 85–130.

The 24.6 point displacement of the relative import price may be questioned. Added to the 53.2 percentage point cumulative displacement of the domestic price level, the implied displacement of the nominal price of imports would be 77.8 percent in dollars or 61.6 percent in foreign currency (77.8 minus 16.2). Achievement of purchasing power parity, as suggested in the last paragraph, would require a displacement of the foreign price level by 37 percent. Thus in foreign currency those foreign goods purchased by the United States would fall in price by 24.6 percent relative to all other foreign goods. Although some raw materials purchased by the United States may have low price elasticities of demand and may exhibit a relative price decline in response to a U.S. recession, the 24.6 percent relative price shift appears implausibly large for U.S. imports taken as a whole. As suggested above in our discussion of table 4, we believe that the large coefficients in the import price equation on the food-energy variable reflect a particular concurrence of events in 1974 that is unlikely to be repeated, and believe that the simulation of the basic model in the first row of table 5 may be too optimistic.

### EXOGENOUS INTERNATIONAL PRICES

The projections given by the dashed lines in figure 1, path II, correspond to the intermediate model of row 3 in table 5, which treats the two relative price variables, $z_3$ and $z_4$, as exogenous, but allows the exchange rate—which exhibits plausible behavior—to remain endogenous. Now there is a greater difference between the unemployment rates in the control and Volcker solutions, with the latter yielding a peak unemployment rate of 8.5 percent in 1984:2. The undiscounted cumulative output loss in the dashed-line projections is double that in path I, and the unemployment rate remains above 7 percent until 1988 instead of 1986. At the end of the simulation the Volcker unemployment rate has reached 5.1 percent and is still falling very rapidly, implying substantial instability after 1992.

### SYMMETRIC VAR MODELS

In contrast to the model described in table 3, which exhibits many empty cells indicating that a particular set of coefficients has been set to zero in a particular equation, the VAR model reported in row 5 of table 5 includes four lagged values in all equations, except that for inflation,

which includes twenty-four lagged values of the dependent variable.[42] Current values are included in the recursive manner of table 1, except that inflation and the output ratio are simultaneously determined. Another difference is the appearance of M1 growth in all equations including that explaining the inflation rate. This VAR system produces a total lack of significance of money changes in the inflation equation: none of the coefficients on current or lagged money is individually significant, even at the 10 percent level, and the $F$-ratio on the inclusion of the current and lagged values is only 0.28. Corresponding to this lack of significance is the identical set of simulation results on rows 5 and 6 of table 5, which respectively include and exclude money from the inflation equation.

More interesting are the much lower sacrifice ratios, both with and without discounting, for the VAR models on rows 5 and 6 as compared to the most closely corresponding restricted model in row 1. The VAR results appear implausible because they imply continuous drifting of real endogenous variables through 1992, even though the growth rate of M1 under both simulations is constant after 1985. For instance, in 1992 under the control simulation the relative price of imports is steadily rising at an annual rate of 4 percent a year with a constant exchange rate, whereas under the Volcker simulation the relative price of imports is rising at the same 4 percent rate but the exchange rate is depreciating steadily at 4 percent a year. By 1992 the *level* of the exchange rate has actually depreciated in the Volcker simulation compared to the control simulation, implying that in the long run restrictive monetary policy *raises* foreign inflation. Further, the cumulative 1981–92 displacement of the relative price of imports is 53.1 percent, which is more than twice as much as in the basic model of the first row in the table and is thus even more implausible than the result that we questioned above.

By making small changes in the VAR model, it is possible to obtain even lower sacrifice ratios. Row 7 shortens the distribution on the lagged dependent variable in the inflation equation from twenty-four to four quarters, thus quickening the overall responsiveness of the model. Then in row 8 the ordering is reversed from that in table 3, with the exchange

42. There are a few remaining asymmetric features of the VAR model in row 5 of table 5 that are necessitated by the limited degrees of freedom. The 1972 and 1973 dummy variables appear only in the exchange rate equation; the Nixon dummy variables appear only in the inflation equation; and, because of its limited 1972–80 sample period, there are only two lags on each variable in the foreign exchange equation.

rate first, the inflation rate next, and so on. This version actually yields a zero discounted sacrifice ratio.

Finally, rows 9 through 11 make a gradual transition to the more conventional VAR models estimated by Sims and others. Row 9 takes the row 7 model and excludes all supply variables except for the relative price of imports (the latter variable is retained because it is used in the six-variable model in Sims' original VAR paper "Macroeconomics and Reality"). This smaller model in row 9 retains the basic properties of row 7, with little change in the discounted sacrifice ratio. But the model of row 9 would never be chosen by a VAR afficionado, since our previous research has been used to introduce the natural output "adjustments" to the $\hat{m}$ and $\hat{Q}$ variables, as well as to state the import price variable in relative rather than nominal form. The last two rows, row 10 in first differences and row 11 in levels, eliminate these adjustments. The model in row 11 seems to us a good example of the folly of the atheoretical VAR approach when it is unencumbered by common sense. The discounted sacrifice ratio is an enormous 34.2, and the implied unemployment rate in the model grows steadily to almost 15 percent by 1992. Why? The specification in levels rather than growth rates mixes up trend and cycle phenomena. It yields a *negative* coefficient on output and a negligible positive coefficient on money in the price equation, which as a result is little more than an autoregression in which the inflation rate responds very sluggishly to restrictive monetary policy.[43]

Overall, we find little to dissuade us from our preference for the basic model. It is based on an inflation equation that is stable over the 1954–80 sample period and in which coefficients have correct signs and are of reasonable size. The auxiliary equations added for the policy simulations yield plausible paths for the endogenous variables, except for the excessive response of the relative price of imports. The version shown in row 2, which restricts the growth rate of the relative import price variable to be zero during the simulation period, omits this implausible import-price pattern and thus seems to us to be the most reliable indication of the consequences of the control and Volcker policies. The VAR models of rows 5 and 6 lack plausibility, since they yield continuous

---

43. The model shown in the last row of table 5 with variables stated as log levels is the same as that in Sims, "Interwar and Postwar," with his interest rate replaced by our import deflator.

long-run drift in real variables many years after the growth rate of M1 in our simulations has arrived at its steady-state value. Finally, we find the VAR models of rows 7 through 11 inferior due to the omission of significant variables.

## Conclusions

This paper has attempted to provide new measures of the output cost of disinflationary monetary policy using traditional and vector autoregressive techniques and to use this substantive issue as an occasion to provide an assessment of alternative econometric methodologies. Our conclusions are divided between those of methodological interest and those that relate to the estimated sacrifice ratios and their policy implications.

### METHODOLOGY

Although to date VAR models have mainly been used for multivariate exogeneity and causality analysis, they also serve in testing the specification of traditional econometric models. Thanks to the discipline imposed by the VAR technique, we have discovered that the relative price of imports, and of food and energy, both usually treated as exogenous, can be partially explained by lagged values of other variables. As a result, the estimated response of inflation to restrictive monetary policy is amplified.

The traditional and VAR approaches can be viewed as selecting different methods of allocating zero restrictions in the face of scarce degrees of freedom. Like any trade-off in economics, the best way to allocate these restrictions should depend on an assessment of benefits and costs. We find that the VAR technique, although a useful tool for checking traditional specifications, has a low benefit-cost ratio. The pursuit of symmetry leads an investigator to omit "special variables" that matter for particular equations such as the effect of the Nixon controls in the inflation equation or the investment tax credit in investment equations. By clinging to published data and eschewing our natural output adjustments, VAR models also tend to mix secular and cyclical effects and to yield biased coefficients for key relations. As an example,

the endogenous treatment of food-energy and import prices suggested by the VAR technique yields an implausibly large response of the latter variable in our simulations.

A VAR enthusiast might be willing to admit that a pure VAR model is of limited usefulness for studying our particular substantive question over a long postsample time horizon and to retreat into a defense of VAR models for multivariate exogeneity and causality testing. But, as the example on row 11 of table 5 illustrates, a VAR model not unlike those published in the literature can yield coefficients that are severely biased and imply a Phillips curve with a perverse slope. This is quite likely to influence the results of exogeneity and causality testing.

### THE OUTPUT COST OF DISINFLATION

The discounted sacrifice ratio that emerges from our basic model is 4.3 with the relative import price variable included and 5.8 with that variable excluded. The latter estimate, which we prefer, suggests that to achieve by restrictive monetary policy a long-run reduction in the inflation rate of 5 percentage points the nation must choose to give up output having a present value of 29 percent of a year's natural GNP, almost $1,000 billion at current prices.

Disinflationary monetary policy in the United States is likely to create similar conditions abroad. Without estimating separate equations for the rest of the world, we cannot conjecture about the size of the additional output lost elsewhere. To the extent that nominal wages and prices are less sticky in other countries, the adjustment process may be less painful there than in the United States. But there is no doubt that the $1,000 billion figure understates the worldwide output loss imposed by the current official monetary policy of the U.S. government.

The output loss from disinflation, however large, does not by itself contain implications for economic policy. The discounted welfare gain from a permanent reduction of the inflation rate by 5 percentage points is unlikely to approach $1,000 billion unless nonneutral tax distortions and financial regulations are assumed to be permanent. We find such a presumption implausible. Further, we believe that the public aversion to inflation largely reflects a confusion between the effects of inflation itself and the real income loss caused by the oil price shocks and productivity slowdown of the 1970s. Economists have a responsibility

to educate the public about the true costs of inflation in a neutral tax and regulatory environment and about the output cost of reducing inflation.

Our paper also has implications for the literature on inflation and Phillips curves. By including the exchange rate and import prices in the U.S. inflation equation, we tie the study of inflation in the United States more closely to the literature on international monetary economics than has traditionally been the case. Just as foreign economists have long recognized, the mix of monetary and fiscal policy, through its effect on the exchange rate, matters for the short-run inflation adjustment process.

Finally, we find the stability of our basic inflation equation before and after 1967 to be encouraging and offer this evidence in rebuttal to those economists who specialize in "sorting through the wreckage" of earlier Phillips curves and prematurely announcing the demise of Keynesian economics.[44]

44. See especially Robert E. Lucas, Jr., and Thomas J. Sargent, "After Keynesian Macroeconomics," in Federal Reserve Bank of Boston, *After the Phillips Curve: Persistence of High Inflation and High Unemployment,* Conference Series 19 (FRBB, 1978), pp. 49–72.

# Discussion

Franco Modigliani found the paper's comparison of the two estimation strategies interesting; but he suggested that the sacrifice ratio implied by the Gordon-King coefficients were not very different from previous results, though were at the very low end of the range. William Fellner criticized the way the authors posed the choice open to policy. They compute the cost of a resolute disinflation policy and compare it with the cost of stabilizing inflation at its present rate. But Fellner argued that holding the present inflation rate is not feasible because it is not a credible policy. If attempted, economic agents will believe that any higher inflation rate will be accommodated. Fellner concluded that any credible policy must involve a commitment to reduce the rate of inflation. Hence he suggested that the costs of wage and price controls would be a more suitable alternative against which to measure the costs of resolute disinflation. Gordon responded that stabilizing inflation at its current level was a feasible policy as long as authorities made clear their determination not to tolerate any further increases in the rate of inflation, including increases caused by random shocks.

Several participants were unconvinced by Gordon and King's attempt to confront the Lucas critique. They reasoned that, in principle at least, the steadfast pursuit of disinflation could convince economic agents that the future would be different from the past and so could reduce the real costs of disinflation.

Christopher Sims asked whether the resolute disinflation policy outlined by the authors was adequately specified. The paper ignores fiscal policy, yet there is some question about whether the resolute disinflation policy pursued by the Federal Reserve, combined with historically high and growing deficits, could result in inflation reductions and output losses consistent with the estimated sacrifice ratios. Sims suggested that the sacrifice ratio of the present policy combination could be much higher than that estimated by Gordon and King.

Sims criticized the authors' comparison between the VAR procedure and the more traditional structural statistical framework. He found the comparison misleading because the VAR models are essentially unre-

stricted while the structural equations take advantage of restrictions based on the authors' claimed a priori knowledge. Sims argued that when VAR models are used for making projections, some procedure for damping sampling variation in estimated coefficients is essential to good performance. A meaningful test of the value of the authors' claimed a priori knowledge would have compared projections from their model with projections from a VAR model estimated with a loose Bayesian prior, not based on a claim to a priori knowledge about specific equations. King interpreted this comment as compatible with the paper's negative verdict on unrestricted VAR models of the type that have recently been popular and welcomed Sims's suggestions for an improved formal methodology for introducing such restrictions.