J. BRADFORD DE LONG
Harvard University

LAWRENCE H. SUMMERS
Harvard University

How Does Macroeconomic Policy Affect Output?

THE NATURAL RATE HYPOTHESIS, with its corollary that demand management policies cannot affect an economy's long-run average level of unemployment or output, has come to be widely accepted even by Keynesian economists. This view is enshrined in standard textbooks: Robert Hall and John Taylor, for example, "stress . . . that macro policy cannot influence the average rate of unemployment. It can only influence the fluctuations of unemployment around the natural rate."[1]

In this paper we raise questions about the validity of the natural rate hypothesis and argue that demand management policies can and do affect not just the variance, but also the mean, of output and unemployment. As a way of comparing the effectiveness of different demand management policies in stabilizing national economies, we return to the much-discussed comparison of macroeconomic performance in the United States and other industrial nations before and after World War II. Previous explorations of macroeconomic performance in historical perspective have focused either on the volatility of output about trends or on the volatility of changes in output.[2] But volatility is not the relevant

We would like to thank Robert Barro, Robert Barsky, Bruce Lehmann, Christina Romer, Andrei Shleifer, Daniel Sichel, Robert Waldmann, members of the Brookings Panel, and participants in seminars at Princeton University and Northwestern University for helpful discussions.

1. Robert E. Hall and John B. Taylor, *Macroeconomics: Theory, Performance, and Policy*, 2nd ed. (W.W. Norton, 1988), p. 486.

2. See, for example, Arthur F. Burns, "Progress toward Economic Stability," *American Economic Review*, vol. 50 (March 1960), pp. 1–19; J. Bradford De Long and Lawrence

measure if, as Keynes and the early Keynesians believed, successful macroeconomic policies fill in troughs without shaving off peaks.

Using the prewar GNP series created by Christina Romer to correct for excess cyclical variability in the standard Kuznets-Kendrick-Gallman series, we find evidence of considerable improvement in U.S. macroeconomic performance since World War II.[3] The average gap between real aggregate demand and the potential supply of the economy has been reduced since the pre-Depression period by an amount that would now run at almost $50 billion a year. This improvement in the performance of the U.S. economy relative to potential is conventionally, and we think correctly, attributed to the more stable financial system, automatic stabilizers, and possibly improved discretionary macroeconomic policies that together make up the postwar Keynesian institutional order and have managed demand for the past 40 years.

H. Summers, "The Changing Cyclical Variability of Economic Activity in the United States," in Robert J. Gordon, ed., *The American Business Cycle: Continuity and Change* (University of Chicago, 1986), pp. 679–733; Charles Schultze, *Other Times, Other Places: Macroeconomic Lessons from U.S. and European History* (Brookings, 1986); Martin N. Baily, "Stabilization Policy and Private Economic Behavior," *BPEA, 1:1978,* pp. 11–60. All focus on the variability of output about trends. Steven M. Sheffrin, "Have Economic Fluctuations Been Dampened? A Look at Evidence Outside the United States," *Journal of Monetary Economics,* vol. 21 (January 1988), pp. 73–84, examines the variance of annual output growth rates and finds no sign of significant stabilization. We find Sheffrin's examination of the variability of growth rates only to be inappropriate because we do not see any close connection between the variability of annual growth rates and economic welfare.

3. The Kuznets-Kendrick-Gallman series is the standard GNP series used in U.S. Department of Commerce, Bureau of the Census, *Historical Statistics of the United States, Colonial Times to 1970* (Government Printing Office, 1975). Discussion of the Romer series can be found in Christina Romer, "Spurious Volatility in Historical Unemployment Data," *Journal of Political Economy,* vol. 94 (February 1986), pp. 1–37; "Is the Stabilization of the Postwar Economy a Figment of the Data?" *American Economic Review,* vol. 76 (June 1986), pp. 314–34; "Gross National Product, 1909–1928: Existing Estimates, New Estimates, and New Interpretations of World War I and Its Aftermath," Working Paper 2187 (National Bureau of Economic Research, 1987); "The Prewar Business Cycle Reconsidered: New Estimates of Gross National Product," *Journal of Political Economy* (forthcoming, 1989). Romer's estimates are in their turn artificially smooth—they omit that part of GNP variability uncorrelated with contemporaneous movements in commodity production. Her work has sparked a debate that includes David Weir, "Unemployment Volatility: A Sensitivity Analysis" (Yale University, 1986); and Nathan Balke and Robert J. Gordon, "The Estimation of Prewar GNP: Methodology and New Evidence," *Journal of Political Economy* (forthcoming, 1989). We use the Romer series because it is the least favorable to our conclusions of the available series.

We first stress that the economy's average level of unemployment is likely to be inefficiently high; an increase in the average level of output is likely to be desirable whether or not the operation of the natural rate property makes it infeasible.We go on to review recent theoretical developments that suggest alternatives to the natural rate hypothesis. Recent years have seen the emergence of a variety of theoretical models that exhibit multiple equilibriums associated with different levels of production. These models, which capture many of the ideas discussed in *The General Theory*,[4] rely on mechanisms like credit failures, low-level demand traps, and asymmetries in price adjustment to generate economies with multiple equilibriums at which the forces pushing for full use of resources are at most very weak. If, as is plausible, demand management policy can affect which of the many possible equilibriums an economy attains, it can have a lasting effect on the level of output. One piece of evidence that suggests that these theories have empirical backing is the asymmetric response of U.S. output to nominal shocks.

Turning to the data, which are discussed and built up in the appendix, we begin by examining the serial correlation of output over the pre-Depression and postwar periods. The time series properties of output have been a major issue in the rapidly growing literature on the presence of "unit roots."[5] Some investigators have concluded that output fluctuations are dominated by permanent shocks—that a 1 percent fall in output this year means that forecasters should revise downward their forecast of output a generation hence by the full 1 percent or even more. This dominance of permanent shocks has been interpreted as revealing that macroeconomic fluctuations arise not from the demand side but from permanent changes in the economy's production technology.[6] We show, following John Cochrane, that substantial persistence in output is largely a postwar phenomenon both in the United States and in the rest of the West.[7] Before the Depression, permanent shocks to output

4. John Maynard Keynes, *The General Theory of Employment, Interest and Money* (Macmillan, 1936).

5. For example, John Y. Campbell and N. Gregory Mankiw, "Are Output Fluctuations Transitory?" *Quarterly Journal of Economics*, vol. 102 (November 1987), pp. 857–80.

6. Matthew D. Shapiro and Mark W. Watson, "Sources of Business Cycle Fluctuations," Working Paper 2589 (NBER, 1988).

7. John Cochrane, "How Large Is the Random Walk in GNP?" *Journal of Political Economy*, vol. 96 (October 1988), pp. 893–920.

accounted for at most a small part of the year-to-year variance in production.

Some of this increase in persistence might arise from an increased variability of potential output and technology growth since the war. More plausible in our view is the hypothesis that most of the shift in the serial correlation properties of output arises from successful demand management policies that have largely eliminated the transitory declines in output, caused by movements to inferior equilibriums, characteristic of the prewar period. That unemployment rates were more skewed before World War II than they have been since lends additional plausibility to this hypothesis.

To estimate the size of the postwar improvement in performance relative to potential, we construct average output gaps by interpolating potential GNP between major cycle peaks. Our output gap measures suggest substantial improvement in performance not only in the United States but also abroad. We also show some empirical support for viewing business cycles as gaps rather than as cycles around supply-driven trends. The existence of cyclical asymmetries, the correlations of constructed gaps with observed unemployment, and the stronger response of output to negative than to positive monetary shocks together suggest that the gaps view may provide a more accurate characterization of fluctuations than does the more standard view of fluctuations as near-symmetric cycles around unique equilibrium trend levels of output and unemployment.

The data analysis in the main body of the paper deals with the pre-Depression and postwar periods. Omitting the largest transitory fall of output below potential biases the case against finding either a significant improvement in performance in the postwar era or evidence that business cycles are best thought of as asymmetric lapses beneath potential, not symmetric fluctuations about trend. In the final section of the paper, we examine the Depression in some detail. The recovery from the Depression carried U.S. output almost all the way back to its pre-Depression trend even before the stimulus of World War II began to affect the level of U.S. production. The level of production reached immediately before World War II suggests that, despite greatly reduced capital investment and a substantial labor force withdrawal, the Depression does not seem to have cast a large forward shadow onto U.S. output. This argues for theories of multiple equilibriums based on considerations other than the

accumulation of human and physical capital.

Cycles vs. Gaps

The proposition that the U.S. economy's equilibrium level of output is inefficiently low is relatively uncontroversial.[8] To begin with, substantial tax rates on labor income create a wedge between the private and social returns to increased employment. In addition, the pervasiveness of monopoly power in the economy creates some presumption that output is below its efficient level. This presumption is reinforced by evidence that the reservation wages of the unemployed are frequently considerably below going wages, and by arguments involving congestion effects as the unemployed search for work.[9]

It follows that increases in the economy's average level of output and employment would almost certainly be beneficial. The extra output generated would almost certainly more than compensate for any extra wear and tear on capital goods. Demand management policies should, therefore, be used to raise the average level of output—if this is feasible.[10]

8. Hall and Taylor, *Macroeconomics,* use a social welfare function that is symmetric in deviations from the natural rate of unemployment. They, however, adopt this primarily for analytical convenience, arguing (pp. 490–91) that "economists have thought less about the costs of episodes when GNP is above potential. The microeconomic argument supporting the idea that costs are important is the following: The extra work effort needed to push GNP above potential is worth more than is the extra GNP. Instead of working as many hours as they do during a boom and consuming and investing the extra output, the public would be better off with less output and more time to spend on their children, their houses, and in recreation." But they too believe that "there is at least a range where a boom is socially beneficial even though it is privately costly to workers to be working longer hours." They attribute this asymmetry to tax distortions that push average labor supply below its optimal level.

9. Arthur M. Okun, "Upward Mobility in a High-Pressure Economy," in Joseph A. Pechman, ed., *Selected Essays of Arthur M. Okun* (MIT Press, 1983), pp. 171–220. See also Arthur M. Okun, *The Political Economy of Prosperity* (Brookings, 1970); *Prices and Quantities: A Macroeconomic Analysis* (Brookings, 1981).

10. Even in new Keynesian models, it may not be. Many menu cost models, for example, have the property that the average level of output is below the optimal level but that all policy can do is iron out fluctuations. See Olivier Jean Blanchard and Nobuhiro Kiyotaki, "Monopolistic Competition and the Effects of Aggregate Demand," *American Economic Review,* vol. 77 (September 1987), pp. 647–66; also Laurence Ball and David Romer, "Are Prices Too Sticky?" Working Paper 2171 (NBER, 1987).

IMPLICATIONS OF NATURAL RATE THEORIES

That the business cycle consists of repeated transient and potentially avoidable lapses from sustainable levels of output is a major piece of the Keynesian view: there is often room for improvement, and good policy aims to fill in troughs without shaving off peaks. This Keynesian view stands in opposition to the natural rate view that the business cycle is due to expectational errors that alternately push the economy above and below its sustainable growth path. This natural rate view implies, even in its variants most hospitable to Keynesian concerns, that the scope for macroeconomic policy to affect welfare is small.

The heart of the natural rate view is the claim that the relation between production and prices is well captured by the stylized Phillips curve relation:

$$\pi_t = \pi_{t-1} + \beta(D_t), \tag{1}$$

where D_t represents the deviation of actual output or employment from a unique equilibrium value determined by tastes and technologies and consistent with steady and anticipated inflation, and where inflation in the preceding year is a good proxy for the anticipated rate of inflation. Expressions like equation 1 may be found in leading macroeconomics textbooks.[11] The role of lagged inflation in that equation might spring from rational expectations in a context where inflation is nearly a random walk, from adaptive expectations, from long-term nominal wage contracting, or from other forms of nominal inertia.[12]

While nominal inertia provides a compelling explanation for unemployment rates that remain above the normal equilibrium level for the length of a business cycle phase, it is much less compelling as an explanation for unemployment rates below the natural level. The standard presumption in economic theories with rigid prices is that rationing takes place on the short side of the market: while it makes sense to say that buyers cannot buy or sellers cannot sell as much as they want at

11. See Stanley Fischer, Rudiger Dornbusch, and Gordon R. Sparks, *Macroeconomics* (McGraw-Hill, 1985); Hall and Taylor, *Macroeconomics;* and Robert J. Gordon, *Macroeconomics* (Little, Brown, 1987).

12. See Olivier Blanchard, "Why Does Money Affect Output?" in B. Friedman and F. Hahn, eds., *Handbook of Monetary Economics* (Amsterdam: North-Holland, 1988).

quoted prices, it makes much less sense to say that buyers and sellers are forced to transact a greater quantity than they want. The standard Keynesian treatment instead assumes that employment is always demand determined, as if firms could expand employment by compelling workers to accept jobs when unexpected increases in prices reduce real wages.[13]

The rationale for this assumption is rarely made explicit. Sometimes reference is made to contracts entitling employers to force overtime on workers, but it is difficult to enforce contracts that call for people to work against their will, and most cyclical employment gains take the form of increases in employment rather than hours. A second suggestion is that employment rises because workers are fooled and do not realize that real wages are lower in booms, but observation suggests that booms cause few regrets: there are few complaints after cyclical expansions by people who wish they had not been fooled into working. In every other part of economics, price rigidities cause too little to be bought or sold. Only in Keynesian macroeconomics do wage and price rigidities lead, half the time, to quantities in excess of their equilibrium level.

The implications of equation 1 for the efficacy of demand management policy may be seen by summing over time and rearranging:

$$\frac{\sum_{t=1}^{T} D_t}{T} = \frac{\pi_T - \pi_0}{\beta T}. \tag{2}$$

Macroeconomic policies that do not raise or lower the inflation rate over a period do not affect the average level of output and employment over that period.[14] Demand management policies mitigate recessions only to the extent that they choke off expansions. Replacing the lagged inflation rate on the right-hand side of equation 1 with a rational expectation of present inflation would generate a similar conclusion. Expectational mistakes will average out to nearly zero—you can't fool all the people

13. See Hall and Taylor, *Macroeconomics*, p. 490. "Instead of working as many hours as they do during a boom . . . the public would be better off with less output and more time to spend . . . in recreation." But the fact that employment is demand determined keeps the public from making this choice.

14. In expectational instead of accelerationist formulations, even policies that do permanently raise the inflation rate do not affect the average level of output unless they are unanticipated.

all the time—and the pattern of policy will have no first-order effect on average production. Even were it desirable to increase production on average, such an increase would not be feasible.

Equation 2 leads immediately to a chain of reasoning that concludes that cyclical unemployment should not be the focus of macroeconomic policy. Policies can do no first-order net good or harm on the output side without permanently raising or lowering the inflation rate. Policymakers could have avoided the Great Depression—or any other recession that is not followed by an immediate regime shift—only at the cost of incurring a higher steady-state level of inflation today. Why, then, should anyone care about cyclical unemployment? Excess unemployment incurred today because of policy "mistakes" allows a larger boom tomorrow. The business cycle produces welfare losses *only* because consumption is not efficiently smoothed across years.

Robert Lucas has argued that, barring fluctuations as great as the Depression, such welfare losses will be small, might be smaller than the losses from the choice of the wrong long-run rate of inflation, and will certainly be far smaller than the losses from policies that retard long-run growth—as long as the premise that fluctuations leave the average level of output unaltered is given.[15] The fact that unemployment falls heavily on a few does not justify the Keynesian position. No matter what policy is adopted, the same unemployment total must be divided across years. If long-term unemployment is viewed as a special problem, it might even be the case that optimal demand policy requires short, sharp recessions and destabilization—it may be better to leave four people unemployed for six months than one person unemployed for two years—not any form of "leaning against the wind."

The view that business cycles are fluctuations about supply-determined trends has placed the Keynesian wing of American macroeconomists in a dilemma. They take stylized equation 1 as a starting point, but their acceptance of it traps them into fighting for the low ground in their running battles with monetarists: since demand management does not affect average employment, the key issue becomes whether demand affects production for one or three periods in models where the length of a period is left unspecified. Concern with the avoidance of excess unemployment as a principal aim of public policy can be supported only

15. Robert Lucas, *Models of Business Cycles* (Basil Blackwell, 1986).

by a framework that drops the belief that the average is the sustainable level of production.

NON–NATURAL RATE THEORIES

The natural rate hypothesis embodies the plausible idea that economies would always operate at a unique natural level of employment and output but for the effects of transitory factors that cause deviations. In new classical theories, these transitory factors are misperceptions by workers and firms of the level of the money stock. In Keynesian theories, it is stickiness in wages and prices that is crucial. Recent work suggests that the conception of economies oscillating around a unique equilibrium may be inappropriate. Instead, an economy's natural rate of output may be like a person's natural state of health: desirable, normal—and better than average.

One alternative to the natural rate theory is a set of theories that link cyclical fluctuations to credit problems.[16] The old metaphor about "pushing on a string" suggests the nature of the asymmetry inherent in the natural rate: banks can either remain healthy or they can fail. If banks fail there are negative macroeconomic ramifications, but there is no corresponding possibility on the positive side. Suppose the health of financial institutions depends on the discounted value of past unexpected changes in the collateral value of the assets backing their portfolios, with the discount factor depending on the ease with which banks can rebuild their real capital and reserves after a shock. In this case, one would expect that negative deviations of actual from expected inflation would have larger quantitative effects than positive deviations of actual from expected inflation. Stabilizing the growth of nominal aggregate demand would then raise the average level of output as well as reducing its variability.[17]

A second alternative is provided by models with what Robert Hall has called "thick-market externalities." In these models, which can be based on search considerations or on increasing returns, the economy

16. See, for example, Bruce C. Greenwald and Joseph E. Stiglitz, "Examining Alternative Macroeconomic Theories," *BPEA, 1:1988*, pp. 207–60.

17. The asymmetric impact of nominal shocks is examined below. For a formal model that carries implications along these lines, see Mark Gertler and R. Glenn Hubbard, "Financial Factors in Business Fluctuations" (Columbia University, 1988).

can settle at any one of multiple equilibriums. These equilibriums often have the property that the optimal one is where the level of production and the rate of resource utilization are highest.[18] Fully satisfactory theories of equilibrium selection in models with multiple equilibriums have yet to be derived. But it is plausible that in the presence of multiple equilibriums purely nominal shocks may have real effects by causing the economy to move between equilibriums. In such a setting, policy can affect the long-run average level of output and employment if policy is used to shock the economy out of unfavorable but not out of favorable equilibriums.[19]

Perhaps the simplest possible model illustrating this point is that of John Bryant.[20] Bryant imagines that production is so interdependent that the economy's output is determined by the minimum effort put forth by any worker. If increases in effort are minimally costly and if increases in output are evenly split among workers, then there is a multiplicity of equilibriums ranging from one in which no one puts forth any effort to one of "full employment," in which everyone puts forth maximum effort. Any equilibrium is sustainable if workers expect it to be sustained. Bryant's model lacks a place for economic policy, but "sunspot" policies that were thought to determine expectations would determine expectations and could shift the economy from one equilibrium to another.

A third alternative that introduces asymmetry is efficiency wage models, which also offer a reason for policy to affect the average level of output.[21] If wages and prices are quick to adjust upwards but slow to adjust downwards, unanticipated increases in nominal demand will do little to expand production, but unanticipated decreases will have a large effect on quantities and a small effect on prices. Consider an unexpected increase in money that starts to reduce unemployment below its equilib-

18. Robert E. Hall, "Comment," *BPEA, 2:1988*, pp. 587–91; Peter A. Diamond, "Aggregate Demand Management in Search Equilibrium," *Journal of Political Economy*, vol. 90 (October 1982), pp. 881–94.

19. For example, Olivier Blanchard and Lawrence H. Summers, "Hysteresis and the European Unemployment Problem," in Stanley Fischer, ed., *NBER Macroeconomics Annual, 1986* (MIT Press, 1986), pp. 15–78.

20. John Bryant, "A Simple Rational Expectations Keynes-Type Model," *Quarterly Journal of Economics*, vol. 98 (August 1983), pp. 525–28.

21. Lawrence H. Summers, "Relative Wages, Efficiency Wages, and Keynesian Unemployment," *American Economic Review*, vol. 78 (May 1988, *Papers and Proceedings, 1987*), pp. 383–88.

rium level. It is obviously in the interest of employed workers for firms to raise their wages. Efficiency wage considerations arising from turnover, morale, recruiting, or effort suggest that it is in firms' interest to raise wages as well. On the other hand, if there is a negative monetary shock, then employed workers have an incentive to fail to recognize what has happened and to resist wage reductions. It is at least plausible that those adjustments that are in the common interest of employers and present employees will occur more quickly than those that are in the interest only of employers.

These considerations all suggest that policy may affect the first as well as the second moment of output. Better policy may be able to fill in troughs without shaving off peaks. At a minimum, these theoretical arguments suggest that the assumption that averages are invariant to demand management policies should not be made casually.

EMPIRICAL EVIDENCE OF ASYMMETRIC RESPONSES

That the above ideas have empirical backing is illustrated by the asymmetric response of the postwar U.S. economy to monetary shocks, as documented by James Cover.[22] Examining quarterly U.S. data since 1948, he found that positive monetary innovations had effects on output that were small and statistically insignificant. By contrast, negative monetary innovations had large and statistically significant effects. Cover's point estimates are that a positive 1 percent innovation in monetary growth leads to a cumulative increase in output of 0.08 percent after three quarters, and that a negative 1 percent innovation in monetary growth leads to a cumulative decrease in output of 2.44 percent after three quarters.

Asymmetric responses to shocks hold for our annual as well as Cover's quarterly data. We have examined the effect of positive and negative shocks on annual average output, estimating the two following systems of equations. Equations 3 and 4 decompose the previous December-to-December change in the money stock into an anticipated component $E_{t-1}\,\Delta M_t$, an unanticipated shock ϵ_t, and a negative unanticipated shock $\epsilon_t^- = \min(0, \epsilon_t)$; they are estimated both with (β_4 set equal

22. James P. Cover, "Asymmetric Effects of Positive and Negative Money-Supply Shocks" (University of Alabama, 1988).

Table 1. Asymmetric Responses of Annual Average Output to Nominal Shocks, United States, before and after World War II

Period	*Coefficients of output equation*[a]						
	Output one year earlier	*Output two years earlier*	*Year*	*Anticipated component*[b]	*Unanticipated shock*[c]	*Extra effect of unanticipated negative shock*[c]	*Standard error*
System of equations 3 and 4 with year-end-to-year-end nominal monetary shocks: coefficients of equation 4							
1. Postwar	1.11	−0.30	0.011	−2.32	0.19	1.04	0.022
	(0.15)	(0.14)	(0.003)	(0.92)	(0.32)	(0.61)	
2. Postwar	0.95	−0.20	0.007	...	0.06	1.41	0.023
	(0.14)	(0.14)	(0.003)		(0.34)	(0.63)	
3. Pre-Depression	0.49	−0.16	0.012	0.30	−0.02	0.55	0.029
	(0.15)	(0.15)	(0.003)	(0.26)	(0.20)	(0.30)	
4. Pre-Depression	0.53	−0.19	0.012	...	0.02	0.48	0.029
	(0.15)	(0.15)	(0.003)		(0.19)	(0.19)	
5. Prewar	0.78	−0.00	0.003	0.51	0.21	0.59	0.037
	(0.14)	(0.14)	(0.001)	(0.19)	(0.21)	(0.33)	
6. Prewar	1.00	−0.22	0.003	...	0.19	0.59	0.040
	(0.12)	(0.12)	(0.001)		(0.22)	(0.35)	
System of equations 5 and 6 with year-to-year nominal GNP shocks: coefficients of equation 6							
7. Postwar	1.33	−0.62	0.000	9.06	0.56	0.24	0.011
	(0.09)	(0.09)	(0.002)	(1.46)	(0.11)	(0.20)	
8. Postwar	0.92	−0.20	0.008	...	0.45	0.49	0.016
	(0.09)	(0.09)	(0.002)		(0.16)	(0.28)	
9. Pre-Depression	0.46	−0.09	0.017	0.14	0.07	0.29	0.025
	(0.14)	(0.14)	(0.002)	(0.27)	(0.15)	(0.20)	
10. Pre-Depression	0.49	−0.12	0.011	...	0.11	0.23	0.025
	(0.13)	(0.13)	(0.002)		(0.13)	(0.17)	
11. Prewar	0.99	−0.16	0.002	0.18	0.39	0.04	0.016
	(0.13)	(0.15)	(0.002)	(0.17)	(0.14)	(0.21)	
12. Prewar	1.08	−0.28	0.002	...	0.41	0.00	0.034
	(0.10)	(0.10)	(0.001)		(0.14)	(0.20)	

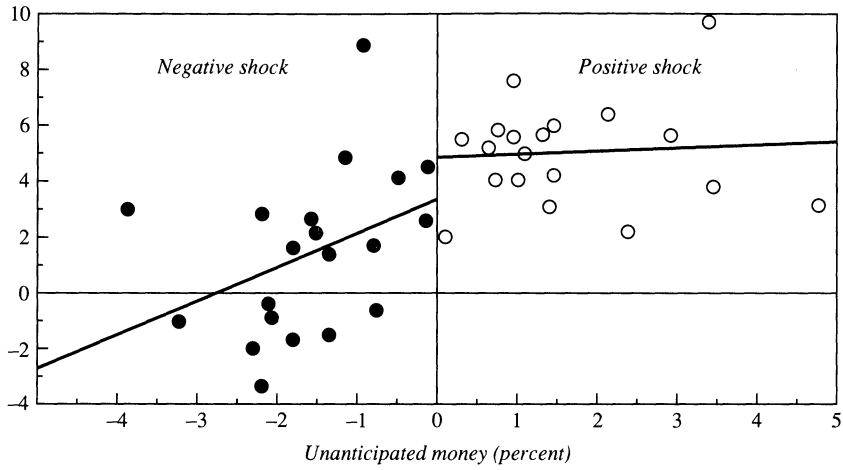
Source: Authors' calculations using the Romer GNP series constructed from Christina D. Romer, "Gross National Product, 1909–1928: Existing Estimates, New Estimates, and New Interpretations of World War I and Its Aftermath," Working Paper 2187 (NBER, 1987); Romer, "The Prewar Business Cycle Reconsidered: New Estimates of Gross National Product," *Journal of Political Economy* (forthcoming, 1989). The Romer GNP series is plotted in figure A-1. Nominal money growth estimates (M2) taken from Milton Friedman and Anna Jacobson Schwartz, *A Monetary History of the United States, 1867–1960* (Princeton University Press, 1963). Nominal GNP estimates obtained by multiplying the Romer real GNP series by the U.S. GDP deflator of Milton Friedman and Anna J. Schwartz, *Monetary Trends in the United States and United Kingdom* (University of Chicago, 1982).

a. Numbers in parentheses are standard errors.

b. Anticipated money growth in system of equations 3 and 4 and anticipated nominal GNP growth in system of equations 5 and 6.

c. Money growth shock in system of equations 3 and 4 and nominal GNP growth shock in system of equations 5 and 6.

Figure 1. Post–World War II Relationship between Unanticipated Money Shocks and Output



Source: Table 1, equation 2.

to zero) and without the restriction that anticipated year-to-year monetary shocks should not have real effects.

$$\Delta \ln M_t = \alpha_1 \Delta \ln M_{t-1} + \alpha_2 \ln Y_{t-1} + \alpha_3 t + \epsilon_t; \tag{3}$$

$$\ln Y_t = \beta_1 \ln Y_{t-1} + \beta_2 \ln Y_{t-2} + \beta_3 t + \beta_4 E_{t-1} \Delta \ln M_t + \beta_5 \epsilon_t + \beta_6 \epsilon_t^- + \eta_t. \tag{4}$$

Equations 5 and 6 examine the extra effect of a negative unanticipated shock $\epsilon_t^- = \min(0, \epsilon_t)$ to nominal GNP growth on real output.

$$\Delta \ln (PY)_t = \alpha_1 \Delta \ln (PY)_{t-1} + \alpha_2 \ln Y_{t-1} + \alpha_3 t + \epsilon_t; \tag{5}$$

$$\ln Y_t = \beta_1 \ln Y_{t-1} + \beta_2 \ln Y_{t-2} + \beta_3 t + \beta_4 \epsilon_{t-1} \Delta \ln (PY)_t + \beta_5 \epsilon_t + \beta_6 \epsilon_t^- + \eta_t. \tag{6}$$

Table 1 reports the results, and figure 1 plots typical values of unanticipated money shocks and output responses. Taking equation 2 in table 1 as typical, we find that a 1 percent positive shock to last December's money stock leads to only a 0.06 percent increase in this year's output in the postwar period. A 1 percent negative innovation in last December's money stock leads to a 1.47 percent decrease in this year's output. The coefficient on negative monetary shocks is signifi-

cantly different from zero. The coefficient on positive monetary shocks is not. And the asymmetric response to the negative innovation ϵ_t^- has a marginal significance level of 0.049 for a one-tailed test. The asymmetry in responses is slightly sharper when the requirement that anticipated money growth be neutral is imposed on equation 4.[23]

Asymmetry is a little stronger before the Great Depression, and is even stronger if the Great Depression is allowed into the sample. By and large, the results are qualitatively similar but weaker when this year's nominal GNP growth rate is the nominal shift variable. The pattern holds for other small changes in specification. Asymmetry is still present when the post-1979 period is omitted from the sample. The possibility that positive monetary shocks proxy, because of the endogenous response of the banking system, for supply shocks that simultaneously raise inflation and the money stock and reduce output also appears to be without foundation: postwar output responds asymmetrically to shocks to real balances as well.

Our annual data thus reveal the same pattern of asymmetric responses in the postwar, prewar, and pre-Depression data that Cover's quarterly data show for the postwar period. In all subperiods, expansionary nominal demand shocks appear to have smaller effects on U.S. output than do contractionary nominal demand shocks—as would be expected if credit-crunch-caused deflations were systematically different from inflations, if rationing on the short side kept output limited to aggregate supply in booms, or simply if the aggregate supply curve were L-shaped. These results suggest that the non–natural rate theories outlined above have empirical relevance. Asymmetry fits more naturally into a framework that sees fluctuations as lapses beneath potential than into one that sees them as cycles around trend. And if fluctuations are lapses beneath potential, then appropriate demand management policies can raise the average level as well as reduce the variance of output.

Comparing Macroeconomic Performance

Both theoretical and empirical considerations thus suggest that the mean level of output is not given by aggregate supply and is not invariant

23. The only peculiarities in the regressions are the large coefficients on the anticipated nominal demand variables in the postwar period. These arise from the productivity

to changes in the pattern of demand management policy. They provide little warrant for accepting as an axiom the proposition that demand management cannot change the average level of resource utilization and output. In this and the following sections, we take as a working hypothesis the proposition that demand management policies have at least the potential to affect the average level of output, and we explore the implications of this hypothesis by comparing the relative macroeconomic performance of the United States and other Western economies before the Great Depression and since World War II.

This comparison is apt because of the dominance from the late 1940s until at least the 1970s of economic policy regimes explicitly aimed at stabilizing demand at a high level, in marked contrast to the regime of "sound finance" that had been in place earlier. The presence in the postwar period of Keynesian institutions and policies has attracted notice from prospective and retrospective observers alike.[24] Deposit insurance and a reinforced commitment by the Federal Reserve to its role of lender of last resort helped stabilize the banking system. The growth of government spending as a share of GNP, coupled with progressive taxation, gave the government budget new weight as an automatic stabilizer. The growth of consumer credit meant that consumer durables purchases were likely to be less adversely affected by income declines.[25] On top of all this came both the recognition by governments that they were responsible for preserving high employment and output and their commitment to fulfill this responsibility.[26]

Whether, and which, elements of this institutional complex played a

---

slowdown and the resulting break in the trend growth of output. The estimated coefficient on the anticipated nominal demand variables is proxying for a break in the time trend.

24. See Burns, "Progress toward Economic Stability"; and De Long and Summers, "Changing Cyclical Variability," for prospective and retrospective views by economists. For a prospective view by conservative and risk-averse securities analysts who did believe that postwar demand management policies would raise output on average, see Benjamin Graham, David Dodd, and Sidney Cottle, *Security Analysis: Principles and Technique,* 4th ed. (McGraw-Hill, 1962), p. 422: "We believe it reasonable to adopt a somewhat more generous approach to the valuation of common stocks than appeared justified in our previous edition. This conclusion is based on the assurance—not formerly present—of massive Federal intervention to prevent a serious business depression. This now appears to be a basic tenet of both political parties."

25. See De Long and Summers, "Changing Cyclical Variability."

26. See James Tobin and Murray Weidenbaum, eds., *Two Revolutions in Economic Policy: The First Economic Reports of Presidents Kennedy and Reagan* (MIT Press, 1988).

significant role in keeping aggregate demand high and stable during the 30 years since the end of World War II can be disputed. We suspect that automatic stabilizers played a much more important role than discretionary fiscal or monetary policy.[27] What cannot be disputed is that total nominal demand has been more stable since World War II than it was before the Depression. The standard deviation of annual nominal GNP growth dropped from 7.0 percent to 3.3 percent.[28] Impressionistic and statistical evidence both point to a reduction in the size of demand shocks affecting the U.S. economy. We attribute this reduction in the volatility of demand to the Keynesian policy regime. Plausible alternative candidates are absent; we see little reason to believe that technology growth or factor prices have been subject to smaller shocks in the postwar period.

The contrast in U.S. economic policy before the Depression and after World War II makes it possible to assess the utility of policies to manage demand. Unless Keynesian policies did improve relative macroeconomic performance in the postwar period, there is little reason to argue for their adoption or continuation. A belief in the utility of Keynesian policies would then be supportable only if the postwar period also saw a striking increase in the natural instability of the economy, and plausible causes of such an increase are difficult to find. Continued belief in the desirability of demand management policies therefore requires not only a conviction that demand management can affect the mean level of output but also a conviction that postwar Keynesian-inspired demand management has raised the mean level of output. We proceed first by examining the time series properties of output, then by assessing the changing gap between actual and potential output, and last by examining the course of the Great Depression.

The Time Series Properties of Output

We begin our analysis by examining the time series behavior of output during the pre-Depression and postwar periods for our sample of countries, detailed in the appendix, for which long-run national account data of acceptable quality exist.

27. See De Long and Summers, "Changing Cyclical Variability."
28. Or 4.8 percent with the World War I period 1915–21 omitted.

THE PERSISTENCE OF OUTPUT

Much has recently been made of the finding, primarily using postwar data, that there exists a sizable unit root in the time series for output. In autoregressive models, coefficients on all lags of output sum to one. Fluctuations in production appear not to die away but to persist indefinitely.[29] One—we believe incorrect—interpretation of this finding is that the presence of a sizable unit root implies that year-to-year fluctuations are either the result of shifts in permanent factors—like tastes and technologies—or the result of transitory, nominal shocks that nevertheless have permanent "hysteresis" effects.[30] Matthew Shapiro and Mark Watson, for example, write as if whether the univariate process for output is highly persistent will determine the choice between "the Keynesian view, in which fluctuations are predominantly transitory, [and] . . . the real business cycle view, in which fluctuations are largely the result of permanent shocks."[31]

An alternative interpretation is that output is the sum of a potential and a cyclical component, with the potential component evolving as a random walk. There is no reason to believe that the potential component of output tends to return to anything that could be called a deterministic trend. Technical progress and capital accumulation do not proceed at constant and deterministic rates.

The business cycle component of output is naturally thought of as stationary over time, as tending to return to some normal level. If the size of the business cycle component shrinks, the potential component may well come to dominate the sample. The persistence of shocks will increase, a unit root will become more readily identifiable, and a

29. Campbell and Mankiw, "Are Output Fluctuations Transitory?" and John Y. Campbell and N. Gregory Mankiw, "Permanent and Transitory Components in Macroeconomic Fluctuations," *American Economic Review,* vol. 77 (May 1987, *Papers and Proceedings, 1986),* pp. 111–17; Charles R. Nelson and Charles I. Plosser, "Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implications," *Journal of Monetary Economics,* vol. 10 (September 1982), pp. 139–62. But see also Peter K. Clark, "The Cyclical Component of U.S. Economic Activity," *Quarterly Journal of Economics,* vol. 102 (November 1987), pp. 797–814; and Mark W. Watson, "Univariate Detrending Methods with Stochastic Trends," *Journal of Monetary Economics,* vol. 18 (July 1986), pp. 49–75.

30. Blanchard and Summers, "Hysteresis and the European Unemployment Problem."

31. Shapiro and Watson, "Sources of Business Cycle Fluctuations."

transitory component may fail to appear in parsimonious estimated models.[32] We see the apparent persistence of postwar output as a sign that the postwar macroeconomic order has performed relatively well and that the business cycle has been small.[33] Our confidence in this interpretation is reinforced by evidence of transitory fluctuations in output before World War II—an era that we believe possessed much larger output gaps, and in which output would naturally possess a much larger transitory business cycle component.[34]

More formally, consider a simple model in which output is composed of a (transitory) gap or cycle and a (persistent) potential or trend:

$$g_t = \gamma + \epsilon_t, \tag{7}$$

where g_t is the current growth rate of potential output;

$$y_t^* = y_{t-1}^* + g_t, \tag{8}$$

where y_t^* is the log of potential output;

$$c_t = \rho c_{t-1} + \eta_t, \tag{9}$$

where c_t is a transitory cyclical component; and

$$\ln y_t = y_t^* - c_t, \tag{10}$$

where $\ln y_t$ is the log of GNP. The change in log output then follows a stationary process:

$$\Delta \ln y_t = \gamma + \epsilon_t + \eta_t + (\rho - 1) \sum_{i=0}^{\infty} \rho^i \eta_{t-1-i}. \tag{11}$$

Writing output in terms of its own past values and of shocks v_t that cannot be forecast from output's past yields:

32. Policy responses to unanticipated shocks will change the persistence of the transitory component and complicate the story. We have argued that automatic stabilizers have been the most effective Keynesian policy devices, and so focus on the case where demand management policy leaves the serial correlation properties of the transitory component more or less unaffected. See De Long and Summers, "Changing Cyclical Variability," where we also argued that effective automatic stabilizers should reduce the persistence of shocks. We now see that argument as misleading; it fails to note that shocks to output in a univariate context include not only business cycle but also long-run productivity components.

33. We pursue this argument at greater length in J. Bradford De Long and Lawrence H. Summers, "On the Existence of a 'Unit Root' in U.S. GNP" (Harvard University, 1988).

34. Cochrane, "How Large Is the Random Walk in GNP?"

$$\ln y_t = (1 + \rho) \ln y_{t-1} - \rho \ln y_{t-2} + v_t - \theta v_{t-1} + (1 - \rho)\gamma, \tag{12}$$

where v_t is a serially uncorrelated shock and $\theta < 1$ is a parameter that together satisfy:

$$(1 + \theta^2)\,\sigma_v^2 = (1 + \rho^2)\,\sigma_\epsilon^2 + 2\sigma^2\eta; \tag{13}$$

$$-\theta\sigma_v^2 = -\rho\sigma_\epsilon^2 - \sigma_\eta^2. \tag{14}$$

A fraction,

$$\frac{1 - \theta}{1 - \rho}, \tag{15}$$

of the shock to output is permanent. As the relative variance of the business cycle component decreases, the degree to which output is persistent—the size of the unit root in output—increases, and θ drops from near one toward zero.

An economist trying to decide whether to model output by equation 12 or by taking output to be a geometric random walk,

$$\ln y_t = \ln y_{t-1} + \gamma + v_t, \tag{16}$$

might well settle on equation 16 and conclude that there are no transitory components even when such components contribute the larger share of the variance. For example, suppose $\rho = 0.75$, $\sigma_\eta = 0.02$, and $\sigma_\epsilon = 0.01$. For these parameter values, the standard deviation of potential output a decade in the future is a reasonable 3 percent, and the transitory component c_t has a standard deviation of 3.2 percent. Then the error variance from fitting equation 16 is only 5 percent greater than the error variance from fitting equation 12. Only after collecting a sample of 135 years would an economist have a 50-50 chance of rejecting equation 16 in favor of the alternative equation 12 at the 0.05 level.[35] Yet the transitory component is responsible for 83 percent of the variance in annual output changes. The implication is that a failure of time series techniques to find transitory fluctuations means not that such do not exist, but only that they do not dominate the data: there are permanent fluctuations mixed

35. If model selection did not give the random walk (equation 16) the advantage of being the null but were instead based on a more balanced goodness-of-fit criterion like the Akaike criterion (see H. Akaike, "Information Theory and an Extension of the Maximum Likelihood Principle," in B. N. Petrov and F. Csáki, eds., *Second International Symposium on Information Theory* [Akadémiai Kiadó, 1973]), then the crossover point would come with a shorter sample. The Akaike criterion would give a 50-50 chance of choosing the true model (equation 12) once the sample amounted to 90 years.

in. A change in institutions, or an increase in automatic stabilizers, that reduces the size of the business cycle may therefore lead to the conclusion that there is no transitory cycle.

The failure to find a transitory cycle using postwar data tells us only that the business cycle component has not been dominant in the sample since World War II, not necessarily that it is nonexistent. Moreover, the prewar period shows strong signs that fluctuations then were as a rule transitory; it is possible to reject the null hypothesis that there was a sizable unit root in output before World War II.[36]

Table 2 regresses output on lagged output and a time trend and provides strong evidence against the claim that output was close to a random walk before the Depression. Both the null and the alternative are unbelievable in that no one thinks either that output is a random walk or that output is a first-order autoregressive process about a linear trend. Table 2 should be thought of as an attempt to quantify the degree to which output is close to a random walk or to transitory fluctuations around a steady trend, not as an attempt to uncover any true generating process.

The presence of substantial persistence in output comes through clearly in the postwar sample. The coefficient on lagged output in the United States is large and only slightly below its expected value if postwar output per person of working age really were generated by a random walk.[37] But the pre-Depression sample rejects with ease the null that the coefficient on lagged output was one in favor of the alternative of first-order autoregressive fluctuations about a linear time trend. The same conclusions hold for other nations. In all except the United Kingdom, the persistence of output rises from before to after World War II.

This conclusion is not sensitive to the yardstick used. An alternative yardstick is provided by the n-period variance ratio, which is the ratio of $(1/n)E(y_{t+n} - y_t - n\gamma)^2$ to $E(y_{t+1} - y_t - \gamma)^2$, where E denotes an

36. Considering the past century as a whole, it is possible to reject the null that output is a random walk in favor of the alternative of an autoregressive process around a linear deterministic trend. See J. Bradford De Long and Lawrence H. Summers, "On the Interpretation of a 'Unit Root' in U.S. GNP," Working Paper 2716 (NBER, 1988).

37. If the null hypothesis that $y_t = 0t + 1.0(y_{t-1}) + \epsilon$ were true, then the estimated coefficient of y_{t-1} is biased and has the Dickey-Fuller distribution calculated by David Dickey. See Wayne A. Fuller, *Introduction to Statistical Time Series* (John Wiley, 1976), p. 373.

Table 2. Output Persistence, Pre-Depression and Post–World War II[a]

Country	*Estimated autoregressive coefficients*[b]	
	Pre-Depression	*Postwar*
United States	0.419	0.816
	(0.150)	(0.081)
Canada	0.621	0.790
	(0.125)	(0.117)
Sweden	0.880	0.951
	(0.095)	(0.106)
United Kingdom	0.671[c]	0.170
	(0.155)	(0.193)
France	0.140[c]	0.792
	(0.217)	(0.139)
Germany	0.049[c]	0.822
	(0.444)	(0.048)

Source: Authors' calculations using the Romer series and data from Angus Maddison, *Phases of Capitalist Development* (Oxford University Press, 1982).

a. Dependent variable is output measured as output per person of working age for the United States and output per capita for all other countries.

b. Value, in each period, of coefficient on output lagged one period.

c. Pre–World War I.

unconditional expectation and γ is the underlying average growth rate of output. The n-period variance ratio is the ratio of the variance of the n-period change in y $(y_t - y_{t-n})$ to n times the variance of the one-period change in y $(y_t - y_{t-1})$. If the underlying series is a random walk, then for all periods n the variance ratio will be equal to one. If the underlying series reverts to a deterministic trend, the variance ratio will approach zero for large periods n. The variance ratio therefore provides an additional measure of the degree to which a series is like a random walk and the degree to which it exhibits mean reversion. Variance ratios adjusted for small-sample bias are reported in table 3, and tell the same story as does table 2.

The clearly transitory nature of fluctuations before World War II is, in our view, enough to remove real business cycle interpretations of postwar output persistence from consideration. Shapiro and Watson's Keynesian–real business cycle menu is more likely a choice between whether demand management policy was deficient or more satisfactory. Something has reduced the relative size of the transitory component in output, and the most plausible candidate is the reduction of the business cycle component of production by the institutions and policies of the

Table 3. International Bias-Adjusted Variance Ratios, Pre-Depression and Post-World War II[a]

Period	*Horizon parameter* 2 years	5 years	10 years	*Horizon parameter* 2 years	5 years	10 years	*Horizon parameter* 2 years	5 years	10 years
	United States			*Sweden*			*Canada*		
Postwar	1.15	1.02	0.87	1.37	1.26	1.28	1.00	0.98	1.30
Pre-Depression	0.91	0.43	0.19	1.06	1.17	0.96	0.79	0.37	0.28
	United Kindgom			*France*			*Germany*		
Postwar	0.79	0.27	0.16	1.19	1.01	0.67	1.51	2.10	3.14
Pre-Depression[b]	1.19	0.82	0.59	0.77	0.41	0.36	0.99	0.49	0.34

Source: Authors' calculations using the Romer series and data from Maddison, *Phases of Capitalist Development.*

a. The n-period variance ratio is the ratio of $(1/n)E(y_{t+n} - y_t - n\gamma)^2$ to $E(y_{t+1} - y_t - \gamma)^2$, where E denotes an unconditional expectation and γ is the underlying average growth rate of output.

b. Pre–World War I.

postwar era. The changing serial correlation pattern of output, in the absence of any strong reason to suspect that the variance of potential growth has become much more unstable, leads us to conclude that there has been an improvement in macroeconomic performance—a reduction in the size of transitory components in output—since World War II.

THE SHAPE OF THE FLUCTUATIONS DISTRIBUTION

Evidence on the serial correlation properties of output does not establish that better policy after World War II has raised the average level of output. It shows only that policy reduced the variance of the asymmetric transitory component in output. One way of exploring this issue is to examine the shape of the distribution of output relative to potential before and after the war. There is some skewness in the distribution of macroeconomic variables. Daniel Sichel has found, using quarterly postwar U.S. data, that the distribution of output relative to trend is significantly asymmetric in levels. Asymmetry appears especially strong when he conditions on the National Bureau of Economic Research business cycle chronology or allows for flexible long-run trends by considering the deviations of macroeconomic variables from cubic splines. Moreover, there is significant asymmetry in industrial production and unemployment rates as well as in output.[38]

38. See Daniel Sichel, "Business Cycle Asymmetry: A Deeper Look," Financial Research Center Memorandum 85 (Princeton University, 1987). Sichel also finds evidence

If postwar policy filled in troughs and shaved off peaks to equal extents, one would expect to see no change in the shape of the distribution of macroeconomic variables. If, on the other hand, policy filled in troughs but did not shave off peaks, one would expect the shape of the output distribution to change between the pre- and postwar periods. The amount of asymmetry should decrease. Since Salih Neftçi and De Long and Summers have found that unemployment is the U.S. macroeconomic variable that shows the most signs of asymmetry in the postwar period,[39] we insure ourselves against the danger of choosing data guaranteed to support our hypothesis by examining the distribution of pre-Depression and postwar unemployment rates for decreases in asymmetry.

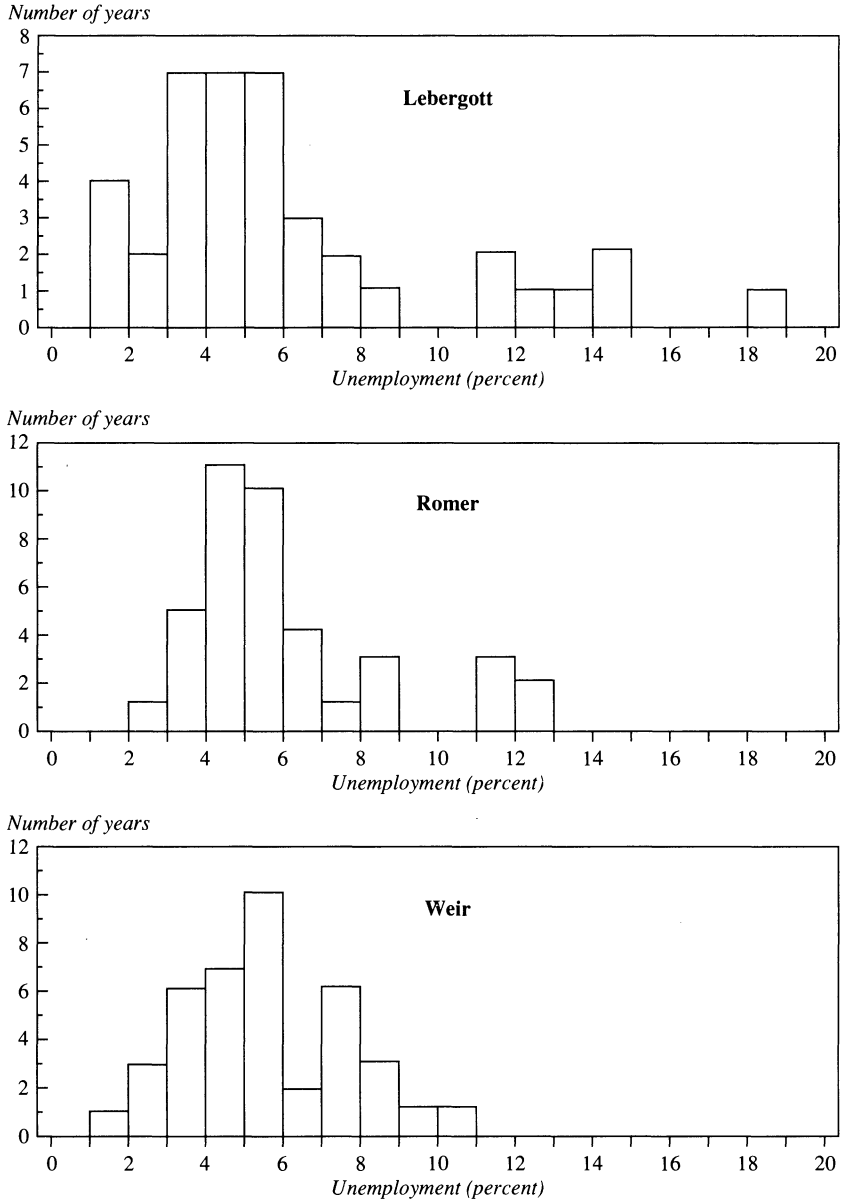
For prewar unemployment rates, we use the original series of Lebergott, the replacement constructed by Romer, and the alternative replacement series constructed by David Weir, who attempts by returning to original sources to overcome the problem of excess volatility originally pointed out by Romer. The nature of unemployment rates forces asymmetry upon them. It is easy to envision unemployment 4 or 5 percentage points above average levels; it is difficult to envision unemployment 4 or 5 percentage points below average. For the pre-Depression period, two of the three unemployment rate estimates exhibit sufficient skewness to reject at the 0.05 level the null of symmetry under the maintained assumption that each year's unemployment rate is an independent normal draw: 1.34 for the Romer series and 1.32 for the Lebergott series, although skewness is only 0.56 for the Weir series. Figure 2 plots the empirical distribution of pre-Depression unemployment rates as estimated by Weir, Lebergott, and Romer. Measured skewness for postwar annual unemployment rates is much less: 0.23 in the raw data, 0.20 in detrended data. Figure 3 plots the Romer empirical distribution of postwar unemployment rates, both raw and detrended to allow for a possible upward drift in measured unemployment for a given degree of labor market tightness in the postwar period.

---

of asymmetry in postwar industrial production data for Canada, France, Japan, and West Germany, but not for Italy or Great Britain.

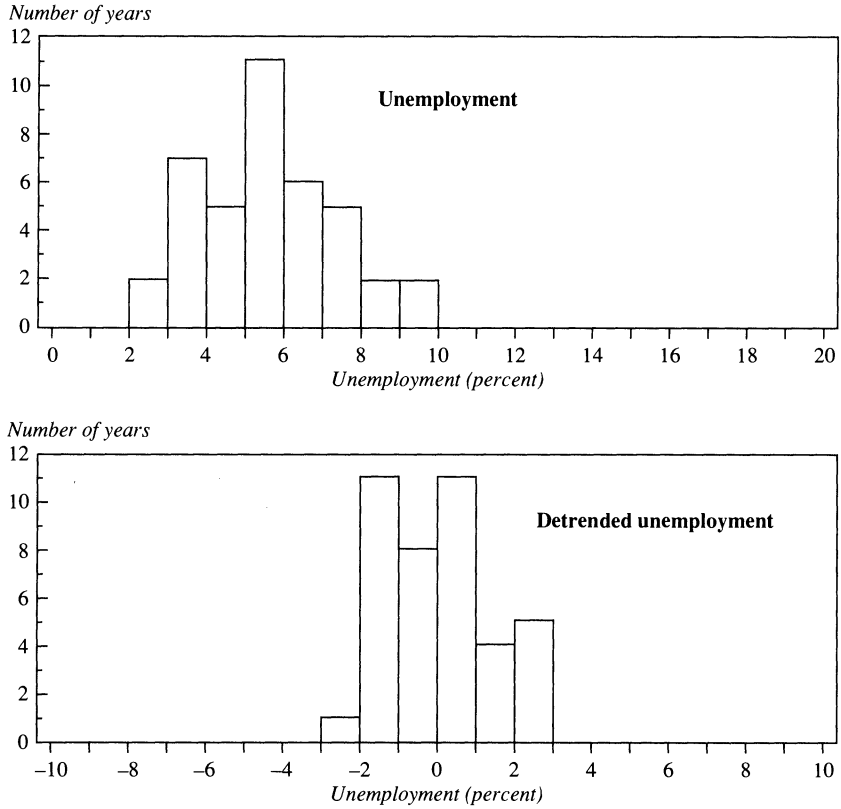
39. De Long and Summers, "Are Business Cycles Symmetrical?" and Salih N. Neftçi, "Are Economic Time Series Asymmetric over the Business Cycle?" *Journal of Political Economy*, vol. 92 (April 1984), pp. 307–28. But Daniel Sichel, "A Reconciliation of Two Empirical Views of Business Cycle Asymmetry," *Journal of Political Economy* (forthcoming), argues on the one hand that Neftçi's evidence is weaker than Neftçi realizes, and on the other that Neftçi's procedures have relatively little power to detect asymmetry.

Figure 2. Histograms of Pre-Depression Unemployment Rates



Sources: Stanley Lebergott, *Manpower in Economic Growth: The American Record since 1800* (McGraw-Hill, 1964); Christina Romer, "Spurious Volatility in Historical Unemployment Data." *Journal of Political Economy*, vol. 94 (February 1986), pp. 1–37; and David Weir, "Unemployment Volatility: A Sensitivity Analysis" (Yale University, 1986).

Figure 3. Histograms of Post–World War II Unemployment Rates



Source: Authors' calculations from post-World War II unemployment data in Romer, "Spurious Volatility."

The decrease in skewness of unemployment rates between the pre-Depression and postwar periods suggests that the reduction in transitory output variability since World War II comes disproportionately from a reduction in the depth of troughs rather than the height of peaks. In the next section we attempt to measure the postwar decrease in the average gap between actual and potential output.

Measuring Output Gaps

If output evolved according to the same stochastic process in the pre-Depression and postwar periods, it would be a simple matter to compare

performance. Any comparison of the variance of innovations, or changes, or deviations about trend would give the same answer. But because both the serial correlation properties of output and the shape of the distribution have changed, the choice of a metric for assessing performance matters. Previous work has concentrated on the variability of output, which will be inappropriate if—as we have argued throughout—demand management policy affects the average level as well as the variance of output.

In this section we therefore revert to a traditional Keynesian approach to the evaluation of performance. We compare the average level of the gap between actual output and potential. We use the economy's peak level of performance to assess potential.

More specifically, we use a simple judgment-free approach to construct a family of potential series. Our approach is motivated by four assumptions:

—First, potential productivity should not decline over time. Workers and managers do not forget production processes.

—Second, actual output is never above potential, except perhaps during total war.

—Third, potential output grows smoothly. New technologies, techniques, and organizations diffuse slowly throughout the economy,[40] as do changes in natural resources, machines, workers, and tastes for work and leisure. Since the determinants of potential change smoothly, potential itself should grow smoothly as well.

—Fourth, actual output does attain potential on a semi-regular basis. Shocks that reduce output below potential are likely to be damped out in a few years.

One could quarrel with these assumptions, but the requirement that output attain potential is close to a definition, and any claim that potential does not grow smoothly because of technology shocks jagged on a year-to-year scale is inconsistent with the picture of technology held by economists who have analyzed technical change.[41] The assumption that potential does not fall is the most vulnerable, for what is an oil or any other negative terms-of-trade shock but a fall in potential? Productive activities that had generated surplus at previous prices no longer do so.

40. See David Landes, *The Unbound Prometheus* (Cambridge University Press, 1967); Alfred Chandler, *The Visible Hand* (Harvard University Press, 1978).

41. Joseph A. Schumpeter, *Business Cycles: A Theoretical, Historical, and Statistical Analysis of the Capitalist Process* (McGraw-Hill, 1939); Simon Kuznets, *Modern Economic Growth: Rate, Structure, and Spread* (Yale University Press, 1966); Nathan Rosenberg, *Technology and American Economic Growth* (Sharpe, 1972).

The assumption requires the further assumption that supply shocks like the oil shock of 1973 are rare and their effects small.

The assumption that output is never above potential except perhaps during total war is defensible. When employment is high, scarce fixed capital is intensively used, learning-by-doing proceeds at a rapid pace, and unemployment rates for marginal groups in the labor force drop to levels not much worse than the core labor force suffers at business cycle troughs. It is difficult to believe that the inefficiencies of too rapid depreciation or too small inventories of unemployed labor are the same order as these benefits of a high-pressure economy.

The assumption that the economy comes close to potential on a semi-regular basis is more vulnerable. It amounts to assuming that the fluctuations of interest to macroeconomics are those that have the duration of the Burns-Mitchell business cycle.[42] The persistence of the Great Depression, and of high unemployment in Europe in the 1980s, indicates that this assumption can be unwise. Indeed, our omission of the 1980s from the sample for this section for European nations is predicated on our inability to construct estimates of potential over the past decade that we regard as reasonable. We think, however, that the assumption is justified for studying the United States in the pre-Depression and postwar periods and for studying European nations in the period before the Depression or World War I and after World War II before the onset of the recent European depression.

The four assumptions inspire the following judgment-free approach for constructing potential output series. Letting y^* denote potential and y denote actual output per person of working age, we construct a family of potential series indexed by k using the following recursive procedure:

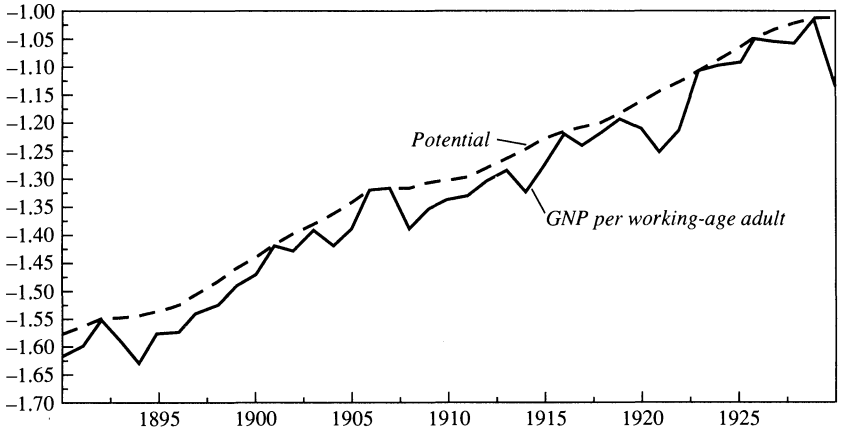
$$y^*_{t+1} = y^*_t + \max\left[0, \max_{i=1 \text{ to } k}\left\{\frac{y_{t+i} - y^*_t}{i}\right\}\right]. \tag{17}$$

The growth rate of potential between year t and year $t + 1$ lies along the straight line with the steepest slope that connects the estimate of potential in year t with actual output per person of working age in any of the years $t + 1$ through $t + k$. The assumption that potential growth is smooth implies that the rate of growth from year t to $t + 1$ will be close to the rate of growth from t to $t + k$. The assumption that actual is never above potential output requires that potential between t and $t + k$ grow at least

42. Arthur F. Burns and Wesley C. Mitchell, *Measuring Business Cycles* (Columbia University Press, 1946).

Figure 4. Peak-to-Peak Potential Output per Working-age Adult, United States, 1890–1930[a]

Log of output per working-age adult



Source: Authors' calculations using the Romer series. See equation 17.
a. Potential GNP in each year based on constructions using GNP of five future periods.

rapidly enough to keep potential at or above actual output. And the assumption that actual attains potential on a semi-regular basis implies that for some year between t and $t + k$ actual output is quite close to potential. Different values of the horizon parameter k correspond to different beliefs about the smoothness of potential growth and the frequency with which actual output draws near to potential. We examine a family of constructed gap series, allowing the horizon parameter k to vary between three and eight years.

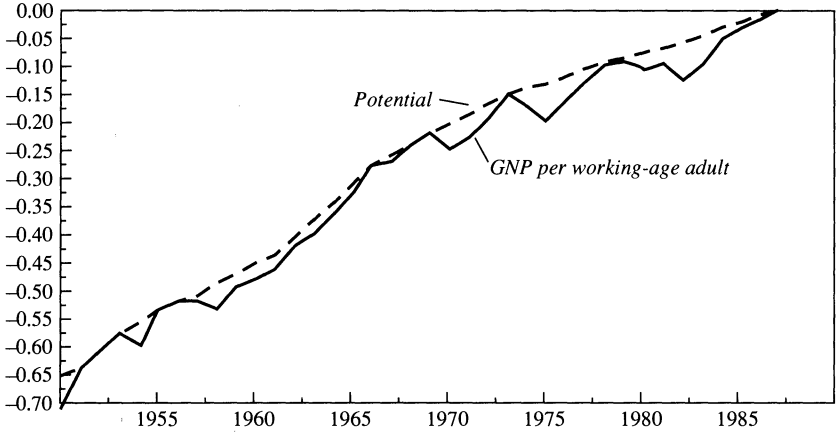
The gap series generated are independent of the average rate of growth of output in the absence of persistent multiyear output declines. But constructed potential output series may well lie below true potential, and estimated average output gaps may be smaller than true average gaps. The average gap measures here should, therefore, be viewed as statistics descriptive of the sample, not as estimates of parameters of underlying generating processes.

CONSTRUCTED POTENTIAL AND OUTPUT GAP SERIES

Generating series for potential output and the output gap for the United States confirms the conclusions about the declining size of the

Figure 5. Peak-to-Peak Potential Output per Working-age Adult, United States, 1950–1990[a]

Log of output per working-age adult



Source: Authors' calculations using the Romer series.
a. Potential GNP in each year based on constructions using GNP of five future periods.

U.S. business cycle inferred from the changing serial correlation pattern of output. For all of the different values of the horizon parameter k used in constructing potential, the mean output gap is at least 50 percent greater before the Depression than after World War II. Figures 4 and 5 present constructed potential based on the Romer estimated GNP series for $k = 5$ in both the pre-Depression and postwar periods. Table 4 presents average gaps for a wide range of horizons k and standard deviations of output from quadratic trends for the Romer series. The mean gap is about 50 percent larger in the pre-Depression than in the postwar period; the macroeconomic slack relative to the economy was some 50 percent greater before the advent of Keynesian demand management policies.

The difference in mean output gaps suggests that the United States has on average come 1 percent of production closer to potential output since World War II than it did before the Depression—a gain that, if sustained, would now run at almost $50 billion a year. Since this improvement in performance is to a first approximation a simple increase in capacity utilization, it is almost all a pure welfare gain. If there is an association between a high-pressure economy on the one hand and rapid growth of potential productivity on the other, then the benefits from

Table 4. Peak-to-Peak Average Output Gaps, United States, Pre-Depression and Post–World War II

Percent

Period	Horizon parameter[a]			*Standard deviation of output from trend*	*Standard deviation of changes in output*
	3 years	*5 years*	*8 years*		
Postwar	1.6	2.3	2.8	2.9	2.6
Pre-Depression	2.6	3.5	4.3	3.4	3.8

Source: Authors' calculations using the Romer series. See equation 17.
a. Number of future periods used in the construction of potential.

improved macroeconomic performance may be substantially understated.

The superiority in estimated performance of the postwar period in the United States does not come equally from all postwar decades. The improvement stems largely from reasonably good performance in the 1950s and excellent performance in the 1960s. Pre-Depression average gaps are not 50 percent but 100 percent larger than gaps during the first half of the postwar period, and the difference between pre-Depression and post-1970 average gaps is only 30 percent.

Such a difference in the size of the average gap is striking given the near-equality of the volatility-based estimates of business cycle size, the standard deviations of levels about quadratic trends, also reported in table 4.[43] In light of the changing persistence properties of output and the large reduction in the mean constructed output gap, we read the near-equality of the standard deviations of output from trend as suggesting that volatility measures have been corrupted by noise introduced by the stochastic nature of long-run potential output growth. The substantial reduction in the volatility of year-to-year changes in output reinforces

43. The standard deviations of growth rates show considerable stabilization, falling from 3.8 percent before the Depression to 2.6 percent after World War II. This disparity in the movement of the variance of levels and differences is a consequence of the changing serial correlation properties of output. Output could become more persistent and yet remain as variable about trends only if the variability of changes were to decline. In a cyclical variability metric, cycles have become longer and less jagged—but not smaller. We find it difficult to argue that such a decline in the variability of annual output growth rates per se is a plus for economic welfare.

Table 5. International Output Gaps, Pre-Depression and Post–World War II
Percent

Period	*Horizon parameter*[a] 3 years	5 years	8 years	*Horizon parameter*[a] 3 years	5 years	8 years	*Horizon parameter*[a] 3 years	5 years	8 years
	United Kingdom			*France*			*Germany*		
Postwar	1.1	1.6	1.9	0.7	0.1	0.1	1.2	1.6	1.9
Pre-Depression[b]	1.9	2.6	0.3	1.9	2.6	2.8	1.1	1.5	1.9
	Sweden			*Canada*					
Postwar	1.0	1.5	1.8	1.2	2.1	2.5			
Pre-Depression	2.2	2.7	3.1	5.0	5.7	6.5			

Source: Authors' calculations using data from Maddison, *Phases of Capitalist Development*. See equation 17.
a. Number of future periods used in the construction of potential.
b. Pre–World War I.

our belief that the small reduction in volatility about trend from the pre-Depression to the postwar period springs from noise introduced into this variability measure by persistent shifts in stochastic potential growth.

Examining the long-run business cycle performance of other Western economies by and large confirms the conclusions we have reached from our examination of the United States. As tables 2 and 3 show, in four of the five other nations the postwar period has seen an increase in the persistence of output fluctuations, suggesting a decline in the relative magnitude of the transitory business cycle component of output. Similarly, as table 5 shows, in four of the five nations postwar output gaps are noticeably reduced.

That measures of volatility about trend in the United States stand alone in opposition to other pieces of evidence—the changing persistence properties of output, the substantial reduction in business cycle size when calculated in a gap metric, and the similar increase in persistence and reduction in business cycle size for other countries—casts doubt on the validity of conclusions that spring from comparisons of U.S. output volatility about trends. Since the revisionist interpretation of relative macroeconomic performance carried out by Romer depends in substantial part on the use of volatility as a proxy for performance, we see no reason to depart from our belief that the postwar period has seen a large improvement in macroeconomic performance.

The difference between the conclusions about the relative sizes of the

pre-Depression and postwar business cycle that emerge from Romer's comparisons of volatility and from our comparisons of gaps demonstrates that whether one believes that business cycles are lapses from potential output or cycles around trend output makes a difference not only for whether one thinks that policy can significantly improve macroeconomic performance but also for whether one thinks that different eras have seen significantly different levels of macroeconomic performance. Cycle-based measures produce little warrant for believing that the Romer series shows that U.S. macroeconomic performance has been significantly better since World War II than it was before the Depression. By contrast, gap-based measures support the claim that the postwar U.S. economy has—for Keynesian or other reasons—achieved superior performance. That the gap-based measures mesh with the conclusions implied by the serial correlation properties of output reinforces what we see as a strong a priori case for gap-based measures.

OKUN'S LAW

Additional empirical evidence exists that the gap approach is superior to the cycle approach. Okun's Law appears stronger in a gap- than in a cycle-based framework, and the distribution of constructed output gaps appears skew in a way that would be expected if the generating process were of the gaps type but is hard to rationalize if it were of the cycles type. These considerations, however, are weak. They may give some added confidence to a reader already sympathetic to this approach; they will not convince those with a strong prior commitment to the proposition that demand management policies do not affect output on average.

Fluctuations in the unemployment rate are an alternative gauge of business cycle fluctuations. To the extent that gaps or cycles really are measures of business cycle components of output, they should be correlated with fluctuations in the unemployment rate. A higher correlation of gap than cycle measures with unemployment would suggest that the constructed output gaps were a better measure of the business cycle than the constructed cycle measures. Indeed, a number of possible unemployment rates appear to be more highly correlated with the constructed output gap than with measures of cycles around trend.

The first part of table 6 compares how the constructed gap ($k = 5$) as a predictor of unemployment fares vis-à-vis output itself and a time

Table 6. Regressions of Unemployment on the Output Gap and on Alternative Measures of the Cycle[a]

Unemployment series	*Output gap*	*Cycle measure*	*Time*	*Rho*	*Standard error of estimate*
Cycle measure: deviation of output from trend					
Weir	−0.564	0.011	−0.036	0.66	1.04
	(0.059)	(0.033)	(0.057)	(0.12)	
Lebergott	−0.846	0.112	−0.191	0.92	1.85
	(0.092)	(0.055)	(0.096)	(0.07)	
Romer	−0.570	0.043	0.038	0.89	1.35
	(0.077)	(0.025)	(0.074)	(0.11)	
Cycle measure: deviation of output from a nine-year moving average					
Weir	−0.543	−0.042	−0.013	0.65	1.02
	(0.095)	(0.072)	(0.004)	(0.13)	
Lebergott	−0.676	−0.071	0.005	0.87	1.63
	(0.151)	(0.115)	(0.007)	(0.11)	
Romer	−0.521	−0.016	−0.000	0.91	1.34
	(0.124)	(0.094)	(0.006)	(0.07)	
Cycle measure: deviation of output from an exponentially smoothed trend					
Weir	−0.621	0.133	−0.021	0.828	0.94
	(0.069)	(0.085)	(0.004)	(0.11)	
Lebergott	−0.738	−0.130	−0.005	0.95	1.59
	(0.116)	(0.143)	(0.007)	(0.108)	
Romer	−0.617	0.209	−0.010	0.98	1.07
	(0.079)	(0.097)	(0.005)	(0.07)	

Source: Authors' calculations using unemployment data from Christina Romer, "Spurious Volatility in Historical Unemployment Data," *Journal of Political Economy*, vol. 94 (February 1986), pp. 1–37; David Weir, "Unemployment Volatility: A Sensitivity Analysis" (Yale University, 1986); Stanley Lebergott, *Manpower in Economic Growth: The American Record since 1800* (McGraw-Hill, 1964); and using the Romer series for output.

a. Entire pre-Depression and postwar sample.

trend. If the proper measure of the state of the business cycle is indeed the deviation of output from a near-deterministic trend, the output gap measure should add no additional explanatory power. Yet the output gap is a better predictor of unemployment.

This result might arise simply from the fact that our potential series captures a relatively flexible trend. Perhaps cycle measures, if allowed a similarly flexible trend, would also perform as well as predictors of unemployment. While we would not be surprised to learn that there is a method of estimating an average trend that does generate a business cycle series as highly correlated with unemployment as the constructed gap, to date we have not found such a method. We have used as cycle measures the deviation of output from long moving averages of output

from exponentially smoothed output (with a smoothing coefficient κ equal to 0.7 that leads to a better fit than the other trend series tried to date), and from a trend series constructed using the nonlinear filter proposed by Robert Hodrick and Edward Prescott.[44] The gap appears a better predictor of the unemployment rate than the cycle-based measures for all three of these proposed flexible average trends.

According to the cycle-trend approach, all years should carry approximately equal amounts of information as to what the sustainable level of production is. In the potential-gap approach, only peak years carry information about the sustainable level of production. The fact that gap-based measures outperform cycle-based measures in predicting unemployment suggests that peak years do carry more information about what the sustainable level of output is than do average years.

The Great Depression

So far we have dealt with the pre-Depression and postwar periods only and have left the Great Depression to one side. In so doing, we have not considered the episode that argues most powerfully both that output fluctuations are transitory and that the average level of output cannot be identified with any supply-driven trend. The Depression is hard to reconcile with the view that the average level of output over time reveals what the sustainable level of output is, for it is next to impossible to make sense of U.S. long-run growth without assuming that the 1929 peak carries a lot more information about what the productive capacity of the economy was than does the 1933 trough. It is impossible to imagine an episode that would push output as much above, as the Depression pushed it below, its average level. The Depression, by its very existence, makes it impossible to argue that fluctuations around averages are symmetrical.

44. Robert Hodrick and Edward Prescott, "Post-War U.S. Business Cycles: An Empirical Investigation" (Carnegie-Mellon University, 1980). The filtered series is affected by the treatment of the endpoints 1889, 1929, 1950, and 1987. We anchor the filtered trend to actual output at the endpoints. The smoothed series thus does not "see" the approaching Depression. If trend did see the Depression in advance, the cycle would be a bad predictor of unemployment.

MEAN REVERSION AND THE GREAT DEPRESSION

Examination of the Great Depression reinforces the belief that output shocks before World War II were transitory. It also reinforces the belief that the economy may possess multiple levels of production that are equilibriums in the sense that the forces pushing output back to potential are weak. Finally, it provides little support for the position that "the time path to equilibrium shapes . . . equilibrium."[45]

The output decline of the Great Depression was not permanent. Output did recover. The log of output per person of working age declined by 0.383—0.455 relative to the 1889–1929 trend—between 1929 and 1933. But between 1933 and 1941 the log of output per person of working age grew by 0.529—0.384 more than the trend. By the time World War II began and the government began to exert command over the economy, more than five-sixths of the Depression decline in output relative to trend had been made up. It is hard to attribute any of the pre-1942 catch-up of the economy to the war. Neither the federal government's fiscal deficit nor the surplus on trade account became an appreciable share of national product before Pearl Harbor.[46]

Figure 6 plots the path of output per person of working age before and during the Great Depression. For reference three trend lines are given. The 1906–29 trend line is the upper linear envelope of the points attained before the Great Depression. The 1892–1929 trend line is meant to capture the long-run average growth rate of output per person of working age. And the 1919–29 trend line is meant to capture any post–World War I acceleration in productivity growth. All three trend lines fall close together. All agree that the degree to which output had recovered from its Depression depths even before the United States entered World War II was substantial.

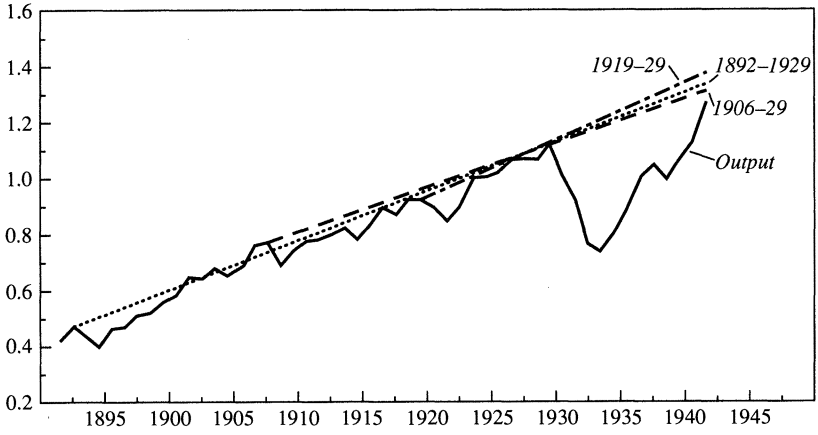
The substantial degree of mean reversion by 1941 is evidence that shocks to output are transitory. Suppose we grant that World War II did not have any significant effect on the U.S. economy until after Pearl Harbor, and assume for the moment that U.S. production really was a

45. Edmund S. Phelps, *Inflation Policy and Unemployment Theory: The Cost-Benefit Approach to Monetary Planning* (Macmillan, 1972), pp. 77–78.

46. Government spending also fails to rise as a proportion of GNP between 1938 and Pearl Harbor.

Figure 6. U.S. Recovery from the Great Depression Measured Against Different Trend Lines, 1890–1950

Log of output per working-age adult



Source: Authors' calculations using Romer series.

random walk: that a 1 percent surprise decline in output led to a 1 percent downward revision of one's estimate of production in the far future. Then the decline and subsequent rapid recovery of the 1929–41 period are orders of magnitude more unlikely than even a decline of the magnitude of the Depression itself.

Calculating the probability of a Great Depression under the null hypothesis that output follows a random walk—a 45.5 percent decline in output followed by a 38.4 percent recovery relative to previous drift within a 12-year period—is straightforward.[47] But such a calculation overstates the evidence against a random walk. Because the contraction was itself unlikely, we divide the probability of such a large decline and recovery by the probability of observing the decline itself. This probability is reported in table 7 for three different estimates of the standard deviation of year-to-year changes in output. It indicates how likely it would be, under the random walk hypothesis, for output to recover to at least 1941 levels given that it declined to at least 1933 levels. The low probabilities of such a recovery argue against the random walk hypothesis and so favor a model with mean reversion.[48]

47. J. Michael Harrison, *Brownian Motion and Stochastic Flow Systems* (Wiley, 1985).

48. The autocorrelations of the annual time series for 1929–39 are different from those of the rest of the prewar period. If the same autocorrelations did prevail, the Depression

Table 7. Probability of Depression and Recovery If Output Follows a Random Walk

Sample period	*Standard deviation of GNP changes*	*Probability of recovery given decline (percent)*
1919–39	7.1	1.06
1889–1933	5.3	0.04
Pre-Depression	3.7	2×10^{-5}

Source: Authors' calculations using Romer series. The probabilities shown are the probability of a decline and recovery as least as great as occurred, divided by the probability of a decline as great as occurred, all under the null hypothesis that output follows a random walk.

Full recovery from the Great Depression weighs strongly in favor of multiple-equilibrium as opposed to hysteresis alternatives to the natural rate hypothesis. Models in which the economy can fall into a low-production, low-activity state and remain there for considerable periods of time with no noticeable tendency to return to equilibrium are consistent with the Depression. Models in which falls in output and increases in unemployment themselves have strong effects on the natural rate of unemployment and natural level of output are not. The Depression as a whole does not seem to have reduced materially the long-run growth path of the U.S. economy.

Full recovery from the Depression is particularly striking given that so many of the mechanisms that economists rely on to produce hysteresis were at work during the Depression. Net capital formation was nil: reproducible tangible assets in 1939 had the same real value as in 1929.[49] Labor force growth slowed: one aim of the Social Security program as instituted was to provide incentives for workers to retire and so reduce excess supply in the labor market. In this aim the program was successful. Labor force participation among men 65 or older, which had drifted downward from 68 percent to 55 percent between 1890 and 1930, had dropped to 41 percent by the 1940 census and continued to fall thereafter.[50] The failure of hysteresis effects to emerge on the aggregate supply side as a result of the Depression suggests that it is appropriate to view

would have been over by 1935: no other business cycle lasted for more than six years. But this is an argument for the line of research followed by Stock that allows each business cycle to evolve on its own time scale, not for any conclusion that declines in output are in general permanent. See James Stock, "Measuring Business Cycle Time," *Journal of Political Economy*, vol. 95 (December 1987), pp. 1240–61.

49. *Historical Statistics*, p. 256.

50. Ibid., p. 132.

the economy in general not as a system in which the path to equilibrium affects the position of equilibrium but as a system with multiple short-run equilibrium positions.

Conclusions

To sum up, we reiterate our four principal conclusions. First, the failure of time series analysis to find strong evidence of a transitory business cycle in the postwar U.S. economy is an argument that recent macroeconomic performance has been good. Before World War II, no one doubts that production contained transitory business cycle components. The Great Depression alone provides sufficient evidence to reject the claim that the canonical shock to production is a permanent one. Changes in potential output are presumably permanent and persistent, while changes in the output gap are likely to be ephemeral and transitory. A reduction in output gaps will change the serial correlation properties of output and may leave time series analysis unable to identify a transitory component. The lack of a transitory component in output since World War II—the finding that the canonical shock to production is persistent in a univariate context—suggests that performance has been good, that shocks that would otherwise have produced severe business cycles have been damped by a robust economic structure or by skillfully conducted policy, and that Romer's findings of little improvement in performance arise because her measures of variability about trends are contaminated by long-run shifts in the stochastic rate of growth of potential output.

Second, assessing performance by gaps establishes a sizable relative improvement in performance since World War II. The significance of this improvement had been cast into doubt by Romer's improved estimates of macroeconomic aggregates, which show only a small decline in volatility between pre-Depression and postwar periods. Romer's work greatly reduces the apparent size of pre-Depression business cycles, but its most extreme interpretation—that there has been no significant decline in the business cycle between the pre-Depression and the postwar period—does not survive the removal of the assumption that demand management cannot affect average output levels and its corollary that volatility is an adequate measure of performance. Similarly, Lucas's argument that demand management cannot significantly improve welfare

rests on the assumption that successful policies cannot raise means but only reduce variances, and also does not survive the shift from a gap- to a cycle-based analytical framework.

Third, there is some theoretical and empirical evidence that weighs on the side of gap- as opposed to cycle-based decompositions of output. Okun's Law is more of a law if output benchmarks are gap-based measures of potential instead of cycle-based measures of average trend. Business cycle asymmetry fits much more easily into a framework in which fluctuations are lapses from full employment than into a framework in which they are cycles about trends. Moreover, present-day microeconomic foundations are to some degree more in sympathy with approaches based on gaps than on cycles: there are plausible mechanisms for generating fluctuations that necessarily cause lapses from full employment rather than cycles about trend.

Fourth, the most important feature of the Great Depression is that it appears to have had few effects on the long-run growth path of the U.S. economy. We read this as a sign that the metaphor of hysteresis as applied to economies should be understood as asserting not that there are no tendencies after demand shocks for the self-regulating mechanisms of the economy to push unemployment down and output back to trend levels, but that there is likely to be a wide range over which the self-regulating mechanisms prove to be weak.

Lucas's and Romer's arguments both carry the implication that Keynesian economics, as practiced in the postwar period, may well have had no significant positive real effects on economic welfare. The core of Lucas's argument is that because Keynesian demand management cannot affect means, it cannot have large welfare effects. An implication of Romer's research is that if Keynesian demand management does not affect means but only volatility about trends, then it has not had positive welfare effects in the postwar period. The natural next step to take along this line of argument—a step explicitly taken by Lucas but not by Romer—is that the unemployment rate should not be an explicit concern of macroeconomic policy. It is plausible to argue that policies aimed at filling in troughs might impart to the economy an inflationary bias that may well have costs. If that is so and if demand management policies cannot provide significant and have not provided positive welfare benefits, macroeconomic policy should ignore indicators of economic slack and instead strive for "sound finance," should attempt to do no

more than provide a stable and predictable environment in which private agents can make their economic decisions, and should leave the unemployment rate to find its own natural level.

This line of argument is compelling if one accepts the premise that demand management policies affect not means but only variances. But we see no reason to have confidence in the correctness of this largely unexamined underlying premise. If the fundamental premise that fluctuations are cycles about means is replaced by the alternative premise—which we find at least as plausible—that fluctuations are primarily lapses beneath sustainable levels of production, then the evidence suggests that largely successful efforts to manage demand have significantly increased average output and reduced average unemployment since World War II.

Moreover, the bias toward inflation contained in postwar demand management policies seems to us relatively small. It is unlikely that the sharp contractions of the prewar period had substantial inflation dividends. We see no sign that booms reached higher peaks before World War II because the ever-present danger of deep depression made the absence of such contractions a large expansionary surprise. And that inflation is less than 5 percent today tells us that the significant reductions in the average output gap since World War II have not been attained at the price of a secular upward drift in average inflation.

The policy implication seems to us to be to stay the course, to keep the unemployment rate an explicit object of policy concern and to keep trying to relieve depressions without damping out expansions. The monetarist policy that the unemployment rate should be left to seek its own level would appear desirable only to an economist dogmatically committed to the belief that demand management cannot affect means. Keynesian demand management policies promise substantial benefits if the gap view fits the world; they carry relatively few costs if the cycle view is correct.

This implication is reinforced by the presence of some theoretical and empirical support for the gap as opposed to the cycle view. Fluctuation-generating mechanisms like those invoked by credit-collapse theories of the business cycle produce depressions but not unsustainable booms. Asymmetry in output and the superior performance of Okun's Law in a gap framework suggest that fluctuations are lapses from potential rather than cycles about trend.

This implication is also reinforced by the bad experiences that have resulted or would have resulted from concentrating macroeconomic policy on maintaining sound finance by following simple policy rules and allowing the unemployment rate to find its own level. In Europe, the highly contractionary demand management policies of the early 1980s have left a legacy of higher unemployment and lower output that has lasted far longer than any believer in the natural rate hypothesis would have predicted and than advocates of such policies implicitly promised a decade ago.[51] In the United States, there is a near consensus that the monetarist policy advocated over 1982–84 of continuing to reduce and stabilize the rate of growth of the money stock and ignoring the then-high unemployment rate would have been a disaster.[52] And the Great Depression itself is interpreted by Milton Friedman and Anna J. Schwartz as the result of the Federal Reserve's pursuit of policies then regarded as sound finance by refusing to inflate radically the high-powered money stock.[53] It may well be that it is Keynesian demand management and its attempt to stabilize demand and employment at a high level that truly provides the most stable environment in which private investors and producers can make their economic plans.

The question of whether Keynesian demand management should continue in the United States turns on whether such policies have achieved significant real output gains over the past 40 years. We note that the United States has suffered only one postwar recession—that of 1982—as severe as those common before the war. We do not see that this avoidance of what in the prewar era would have been seen as the

51. A plausible case can be made that many of the structural problems Europe faces today are in fact the consequences of the dramatic increases in unemployment that occurred during the disinflation of the 1980s. Had the disinflation proceeded more gradually, or had inflation been permitted to remain steady, output and employment might well be much higher today. The contraction of the early 1980s appears to have knocked European output down to an unfavorable real equilibrium, and in the absence of rapid demand expansion there appears to be no strong push back toward full employment, just as there were no strong forces pushing Britain to full employment in the 1920s or the United States to full employment in the 1930s.

52. See Benjamin Friedman, "Lessons on Monetary Policy from the 1980s," *Journal of Economic Perspectives*, vol. 2 (Summer 1988), pp. 51–72; William Poole, "Monetary Policy Lessons of Recent Inflation and Disinflation," *Journal of Economic Perspectives*, vol. 2 (Summer 1988), pp. 73–100.

53. Milton Friedman and Anna Jacobson Schwartz, *A Monetary History of the United States, 1867–1960* (Princeton University Press, 1963).

inevitable share of severe recessions has lowered the output reached in booms. And inflation now is essentially what it was immediately before the Korean War. The United States has successfully avoided all but one (1982) of the postwar equivalents of the 1894, 1908, and 1921 troughs, and yet has not avoided the postwar equivalents of the 1901, 1906, and 1929 peaks. On balance we see postwar demand management as a significant success. The implicit relative odds that most economists less committed to the gap view than we are would give on the truth of the gap and cycle views might well make the continuation of Keynesian demand management a favorable gamble. Whether still further significant improvement would have been possible in the postwar era—whether demand management ought to have been even more aggressively expansionary than it has been on average—is a much more uncertain prospect.

APPENDIX

Data Sources

THE DATA for U.S. output per person of working age used in this paper come from recent reworking of pre-Depression output levels by Christina Romer, deflated by total U.S. population between 16 and 65 taken from *Historical Statistics of the United States*.[54] Nathan Balke and Robert Gordon have also undertaken to review the previous estimates of GNP. Throughout the paper we rely on the Romer series because it is the least favorable to our conclusions. It shows less of a reduction in the size of the business cycle since World War II than does the Balke-Gordon series or the standard series.[55] As table A-1 shows, both sets of data attribute smaller recessions to the pre-Depression economy than do the original estimates of Kuznets. Figures A-1 and A-2 plot the Romer, Balke and Gordon, and standard series for GNP per person of working age over the past century.

The Romer series is biased against finding a significant prewar business cycle for yet another reason. It is a series of fitted values adjusted to

54. Romer, "The Prewar Business Cycle Reconsidered"; and "Gross National Product, 1909–1928."

55. Balke and Gordon, "Estimation of Prewar GNP: Methodology and New Evidence;" *Historical Statistics*.

Table A-1. Peak Declines in GNP per Working-age Adult during Recessions, United States, 1889–1929

Percent change

	Output series		
Recession	*Romer*	*Balke-Gordon*	*Standard*
1892–94	−8.0	−7.9	−10.4
1895–96	. . .	−4.7	−8.7
1901–02	−0.7	−0.5	. . .
1906–08	−7.4	−11.9	−10.4
1913–15	−4.3	−9.9	−10.4
1918–21	−3.8	−10.0	−4.5
1926–28	−1.3	−1.1	−1.1

Source: Romer series from "Gross National Product, 1909–1928"; "The Prewar Business Cycle Reconsidered"; and U.S. Department of Commerce, Bureau of the Census, *Historical Statistics of the United States, Colonial Times to 1970* (GPO, 1975). Balke-Gordon series from Nathan Balke and Robert J. Gordon, "The Estimation of Prewar GNP: Methodology and New Evidence," *Journal of Political Economy* (forthcoming, 1989). Standard series from *Historical Statistics*.

match the standard Kuznets-Kendrick-Gallman series in benchmark years. As a result, it omits any short-run variance in national product not correlated with contemporaneous movements in commodity production. It understates the transitory business cycle component of national product. As Romer acknowledges, in the postwar era the components that are omitted from the prewar series account for about a fifth of output variance.[56]

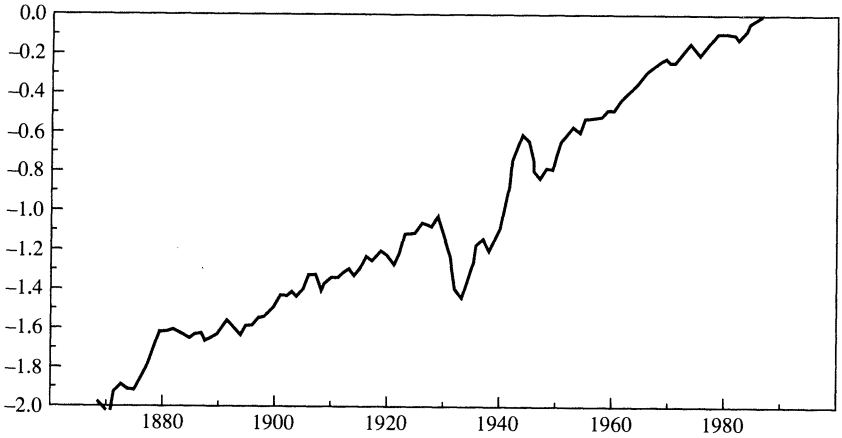
Two points from figures A-1 and A-2 are worth noting. The first is the change in the status of the Great Depression as one moves from the Kuznets to the Balke-Gordon to the Romer estimates. In the standard series, the Great Depression is simply the largest of a number of large prewar recessions. The decline in output per person of working age over the 1921 recession is at least two-thirds, and the declines over the 1893 and 1908 recessions at least one-third, that of the Great Depression. According to the standard series, the Great Depression is but the most severe episode of a disease—severe depression—endemic in the prewar United States. By contrast, the Romer series contains no other recession even one-quarter as bad as the Great Depression.

The second point worth noting is the break in the apparent cyclical pattern before 1890. The decade of the 1880s sees neither large recessions nor labor productivity growth. The change in the apparent character of

56. Romer, "The Prewar Business Cycle Reconsidered."

Figure A-1. Romer Estimates of U.S. Output per Working-age Adult, 1869–1987

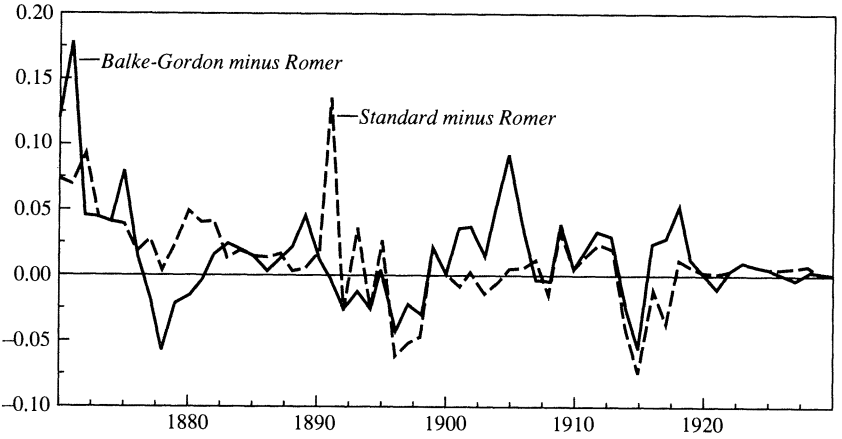
Log of output per working-age adult



Sources: Constructed from data in Christina Romer, "The Prewar Business Cycle Reconsidered: New Estimates of Gross National Product," *Journal of Political Economy* (forthcoming, 1989); and "Gross National Product, 1909–1928: Existing Estimates, New Estimates, and New Interpretations of World War I and Its Aftermath," Working Paper 2187 (NBER, 1987); and NIPA.

Figure A-2. Cyclical Divergence of Standard, Romer, and Balke-Gordon Estimates of U.S. Output, 1870–1930

Log of output per working age adult



Sources: Romer series from figure A-1 and also Nathan Balke and Robert J. Gordon, "The Estimation of Prewar GNP: Methodology and New Evidence," *Journal of Political Economy* (forthcoming, 1989).

the business cycle raises the suspicion that the data before 1889 are of much lower quality.[57] Accordingly, we focus our analysis on the years since 1889.

The unemployment rate data used in subsequent sections consist of the original rate constructed by Lebergott, and the alternatives constructed by Weir and Romer.[58] Weir's and Romer's estimates are almost surely superior to those of Lebergott, which exhibit excess cyclical volatility. The Romer series uses the Lebergott series as raw material and attempts to compensate for excess volatility. Weir tries to build his series up from the raw data while remaining sensitive to assumptions that might introduce excess volatility. Curiously, the Romer output series is more highly correlated with the Weir than with any other unemployment series.[59]

SAMPLE PERIODS

For the United States we consider two sample periods, one post–World War II and one pre-Depression. The postwar period is 1947–87; the pre-Depression period is 1889–1929. We exclude the World War II period from consideration because we believe the mechanisms then at work to produce high levels of output reveal little about the peacetime potential of the economy. We also exclude the Depression.

The exclusion of the Depression biases analysis against finding any significant improvement in economic performance. This may not be appropriate. Perhaps the Depression should be seen as a product of the same economic structure that produced the rest of prewar cycles, in which case the fact that the prewar structure could and did generate such a depression is important evidence of deficiency from the standpoint of macroeconomic performance. Since the Depression is the most virulent outbreak of depression, there may be much to be learned from its study.

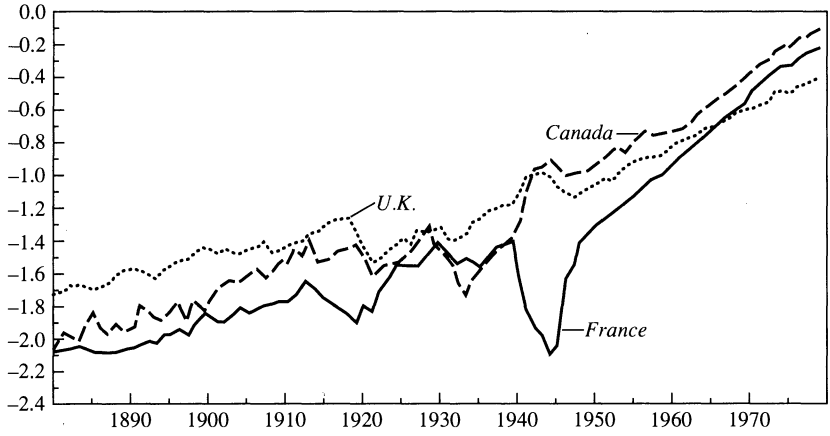
57. See William Shaw, *Commodity Output in the United States* (Columbia University Press, 1947).

58. Stanley Lebergott, *Manpower in Economic Growth: The American Record since 1800* (McGraw-Hill, 1964); Weir, "Unemployment Volatility"; Romer, "Spurious Volatility in Historical Unemployment Data."

59. For the 1889–1929 sample, the R^2 from regressing unemployment on the Romer output series and a time trend is 0.76 for the Weir unemployment series, 0.70 for the Lebergott unemployment series, and 0.55 for the Romer unemployment series.

Figure A-3. National Product per Capita, Canada, France, United Kingdom, 1880–1980

Log of output per capita



Source: Angus Maddison, *Phases of Capitalist Development* (Oxford University Press, 1982).

On the other hand, the Depression is so extraordinary that it dominates statistics over any period in which it is included. We exclude the Depression from the bulk of the analysis, but we return to it and consider the lessons of the Depression at the end of the paper.

INTERNATIONAL EVIDENCE

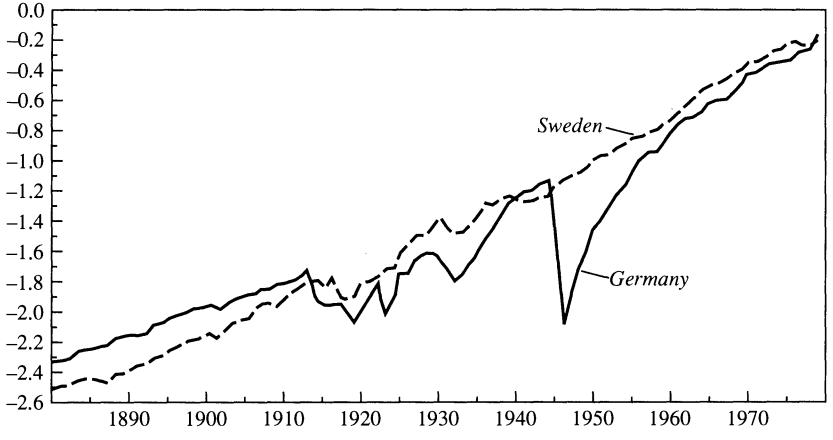
Historians constructing long-run national accounts have quite rightly focused on getting the long-run trend right. They have placed first priority on ensuring that the constructed series provide a good picture of the long-run sweep of growth. Data that are useful for examining not only long-run growth issues but also the business cycle are rare. In our estimation, five industrial economies offer data of reliable quality: Canada, France, Germany, Sweden, and the United Kingdom. Long-run GNP growth data for these five (and for eleven other industrial nations, including the United States) have been constructed and compiled by Angus Maddison.[60] National product per capita for these five nations is plotted in figures A-3 and A-4.

The United Kingdom estimates, derived by C. H. Feinstein, are based on independent expenditure and income series, and are thus probably

60. Angus Maddison, *Phases of Capitalist Development* (Oxford University Press, 1982).

Figure A-4. National Product per Capita, Germany and Sweden, 1880–1980

Log of output per capita



Source: Maddison, *Phases of Capitalist Development.*

the most reliable of any of the historical data.[61] Feinstein was able to use two independent sources of information on aggregate production. The existence of both income-based and output-based measures of total economic activity made his task easier than that of Kuznets for the United States. His estimates are correspondingly less likely to suffer from the kind of excess volatility identified by Romer.

Estimates of French production were constructed by Angus Maddison from data in Carré, Dubois, and Malinvaud.[62] The French estimates appear, if anything, to underestimate cyclical volatility. De Long's examination of the path of output during the post–World War I returns to the gold standard suggest that they understated volatility during the 1920s.[63]

Canadian product estimates are derived from O. J. Firestone.[64] The Firestone estimates for Canada may suffer from the excess cyclical

61. C. H. Feinstein, *National Income, Expenditure, and Output in the United Kingdom, 1855–1965* (Cambridge University Press, 1972).

62. Jean-Jacques Carré, Pierre Dubois, and Edmond Malinvaud, *La Crossiance Française* (Seuil, 1972).

63. J. Bradford De Long, "Returning to the Gold Standard: Britain and France in the 1920s" (Ph.D. dissertation, Harvard University, 1987).

64. O. J. Firestone, *Canada's Economic Development: 1867–1953* (Bowes and Bowes, 1958).

volatility diagnosed by Romer in the Kuznets-Kendrick-Gallman estimates for the United States. The results for Canada should be taken with a grain of salt. Moreover, Canadian macroeconomic fluctuations are closely tied to U.S. fluctuations, and it is questionable that adding Canada introduces much additional real information.

Swedish data are derived from Olle Krantz and Carl-Axel Nilsson.[65] National product estimates for Sweden at the end of the nineteenth century are thought to be of very good quality for a country as poor as Sweden then was. The precocious development of the Swedish government means that a historian of Sweden has more government data on production than do those of most other countries. The Swedish data are unlikely to suffer from excess cyclical variability; widespread use of interpolation is more likely to make Swedish data before World War II too smooth.

The German data are the most questionable. They were constructed by Maddison, and rely heavily on Hoffman's industrial production estimates.[66] The German national product data are sufficiently less volatile than Hoffman's industrial production data to raise the suspicion that the German data may suffer from deficient volatility.

Two different prewar periods are used: a pre–World War I period for the United Kingdom, France, and Germany, all heavily involved in World War I, and the standard pre-Depression period for Canada and Sweden. For all five nations the post–World War II period is taken to be 1948–79. It thus omits the post-1980 period of high unemployment, which we see as involving a shortfall of output from potential of a magnitude unseen since the Great Depression. This omission is partly due to necessity, for our procedures have no purchase on what potential output is in Europe today. This omission also has a substantive rationale. Out of fear of the consequences of real wage rigidity, European governments in the 1980s have abandoned any attempt to stabilize demand at a high level. The prewar to 1948–79 comparison is thus between performance before the Depression and performance while governments were pursuing full-employment policies.

65. Olle Krantz and Carl-Axel Nilsson, *Swedish National Product 1861–1970: New Aspects on Methods and Measurement* (CWK Gleerup, 1975).

66. Maddison, *Phases of Capitalist Development*, p. 164.

Comments and Discussion

N. Gregory Mankiw: Reading this paper brought me back to my days in graduate school when my friends and I were looking for dissertation topics. I remember one of our elders giving us the following piece of advice: "It is not sufficient," he told us, "to choose a question that is interesting and important. You must also choose a question that you have some hope of answering."

Bradford De Long and Lawrence Summers have boldly chosen not to follow that advice. Their question—can demand management policy affect the average level of economic activity?—is obviously interesting and important. And even though they seem to have no way of answering it, at least not convincingly, the paper usefully draws our attention to the issues.

The natural rate hypothesis, the target of this paper, has a prominent place in current macroeconomics. Most economists schooled since the early 1970s accept it as a basic tenet. De Long and Summers persuasively argue that the issue is not really resolved, either theoretically or empirically, and that macroeconomists should not so readily turn to the natural rate hypothesis when thinking about macroeconomic policy.

In assessing the validity of the natural rate hypothesis, we must first ask exactly what the hypothesis is. De Long and Summers do not give us a precise statement of the hypothesis they are suggesting we reject. So let me start by trying to do so.

First, I believe it is useful to make a distinction that De Long and Summers do not emphasize: between monetary policy and other sorts of government policy. The natural rate hypothesis is a claim about the limited ability of monetary policy. Although it may also have implications for fiscal policy, it is not in essence a hypothesis about fiscal policy.

Fiscal policy clearly does affect the mean level of economic activity: taxes discourage market activity, and public goods like highways encourage it. Similarly, regulatory policies, such as the antitrust laws or the Federal Deposit Insurance Corporation, affect mean output to the extent that they correct or create market imperfections. It is only monetary policy to which classical economics attributes neutrality. The natural rate hypothesis, which asserts that classical economics is right in the long run, is thus a hypothesis about the effects of money.[1]

As a statement about monetary policy, the natural rate hypothesis can take a variety of forms. A weak form might be the following: the mean level of economic activity is independent of the mean rate of money growth. This is, I suspect, what Friedman and Phelps had in mind when they proposed the natural rate hypothesis, since the issue of the day was whether the long-run Phillips curve was vertical. Of course, one can think of many reasons that even this weak form of the hypothesis might be false. For example, higher inflation increases shoeleather costs and menu costs and crowds in capital through the Mundell-Tobin effect. Yet most economists judge these effects to be quantitatively unimportant (except in hyperinflating economies), and I suspect they are right. The weak form of the natural rate hypothesis seems a good first approximation, and De Long and Summers do not seem to suggest otherwise.

Instead, De Long and Summers argue against a stronger form of the natural rate hypothesis, which might be stated as follows: the mean level of economic activity is independent of the conduct of monetary policy. This strong form of the hypothesis would be violated if systematic monetary policy, such as offsetting exogenous shocks to aggregate demand, were able to alter the mean level of output or unemployment.

De Long and Summers correctly point out that some textbook models of economic fluctuations exhibit this strong form of the natural rate hypothesis. In these models, monetary policy can affect the variability of output, but not the mean. They also point out that in a variety of more sophisticated models, this strong form of the natural rate hypothesis is violated.

As a theoretical matter, it would be surprising if the strong form of the natural rate hypothesis did hold. Uncertainty plays a central role in

1. When Milton Friedman proposed the natural rate hypothesis, he did it in a section entitled "What Monetary Policy Cannot Do" in a paper called "The Role of Monetary Policy," *American Economic Review*, vol. 58 (March 1968), pp. 1–17.

all kinds of economic behavior. If monetary policy can influence the variability of economic activity, it can surely influence the level as well. For example, if systematic monetary policy reduces the variability of real GNP, it makes the United States a safer place to invest, which induces capital inflows from abroad. At the same time, it also decreases the need for precautionary saving. It is not hard to think of numerous channels through which a reduction in the variability of output alters its mean level.

Even in textbook models, it is easy to find reasons to reject this strong form of the natural rate hypothesis. Many textbooks present the reverse-L-shaped aggregate supply curve. Because of capacity constraints, increases in aggregate demand raise prices more quickly than decreases in aggregate demand lower them. This aggregate supply curve, or indeed any convex aggregate supply curve, will imply that stabilization increases mean output.

This convexity of aggregate supply is potentially important. For example, when economists estimate Phillips curves, the convexity of aggregate supply often enters because the reciprocal of the unemployment rate, rather than the level, enters the equation. Such an expectations-augmented Phillips curve would be

$$\pi = E\pi + \alpha + \beta(1/U).$$

Since expectation errors must average to zero, this specification implies that the Federal Reserve cannot influence the mean of the reciprocal of the unemployment rate. Yet, because of Jensen's inequality, the Federal Reserve can influence the mean of the level of the unemployment rate. Straightforward calculations using unemployment data show that if the reciprocal of the unemployment rate had been stabilized around its mean over the past 10 years, the mean unemployment rate would have been 7.2 percent rather than 7.4 percent. Assuming an Okun's Law coefficient of 2.5, this 0.2 reduction in mean unemployment over a decade is equivalent to 5 percent of one year's GNP. Hence, stabilizing unemployment over a decade is roughly equivalent to averting a moderate-sized single-year recession.

Such a calculation is, of course, merely speculative. As far as I know, there has been little work aimed at examining the convexity of the aggregate supply curve. Perhaps in the future there should be, for it seems that the benefit of economic stabilization depends crucially on the degree of convexity.

De Long and Summers's rejection of the strong form of the natural rate hypothesis is, however, not based on the estimation of such a structural equation. Their reticence to perform any sort of structural estimation seems to stem from an agnosticism over precisely why the natural rate hypothesis fails. The closest they come to structural estimation—and this is my favorite part of the paper—is examining the differential impact of positive and negative demand shocks on real GNP, following work by Cover. Unfortunately, the evidence is not clear-cut: while the point estimates suggest some asymmetry, they are usually not statistically significant.

Most of De Long and Summers's inferences come from examining the univariate properties of real GNP in the prewar and postwar economies. They present us with some intriguing observations. Unfortunately, the relation of these observations to the hypothesis under question is informal and often hard to follow.

For example, one of their observations is that fluctuations in output are more persistent in the postwar period than in the prewar period. De Long and Summers interpret this fact as evidence that macroeconomic policy has improved. One could just as plausibly draw the opposite inference: one might suppose that good policymakers correct their mistakes while bad policymakers let their mistakes persist.

Yet even if we accept the interpretation of the data suggested by De Long and Summers, I am not sure what it tells us about the natural rate hypothesis. There were many changes in the economy between these two periods. De Long and Summers provide no evidence that the observed changes in output had anything to do with monetary policy, or that these changes are informative regarding the effects of monetary policy. Even if we concede that macroeconomic performance improved, and concede that policy of some sort was responsible, I am not sure how to make the leap to rejecting the natural rate hypothesis.

The univariate approach taken by De Long and Summers simply lacks sufficient power to be useful in answering their question. The fact that macroeconomists are rarely in consensus shows how difficult it is to test competing theories in macroeconomics, even when given all the data at our disposal. Examining the time series properties of real GNP alone may refine our set of stylized facts about the business cycle, but it will inevitably fail to produce compelling evidence for or against a claim such as the natural rate hypothesis.

In summary, I believe De Long and Summers have usefully raised some interesting and important questions. I doubt that this paper will convert believers in the natural rate hypothesis. But it should make them more open to the alternatives.

Christina D. Romer: In their paper, De Long and Summers use various types of evidence to argue that macroeconomic performance has improved since World War II. They state that changes in the output gap, the skewness of unemployment, and the persistence of real GNP all suggest that stabilization policy has been successful. In these comments I suggest that some of De Long and Summers's empirical results are flawed and that their interpretation of these results is potentially misleading.

My first comment concerns De Long and Summers's interpretation of their findings about output gaps. They argue that in assessing macroeconomic performance, large falls in production should be counted differently from large rises in production. That is, they suggest that it is more instructive to compare the average deviation of output from potential—the gap—between the prewar and postwar eras than to look at how a simple measure of volatility, such as the standard deviation of percentage changes, changes over the same period. This point is sensible, and if gap- and volatility-based measures of performance yielded different conclusions about stabilization over time, I would find this result very interesting.

From the tone of the paper, one gets the impression that gap-based measures of macroeconomic performance do indeed show much more stabilization over time than do volatility-based measures calculated using the same data. To quote De Long and Summers, "Such a difference in the size of the average gap is striking given the near-equality of the volatility-based estimates of business cycle size." In truth, however, the results that they get using a gap-based measure of the change in macroeconomic performance are very similar to those derived using simple volatility-based measures of performance. This similarity can be seen in table 4 of their paper. Based on my prewar estimates of GNP, the ratio of the standard deviation of percentage changes in the pre-Depression era to that in the post-World War II era is 1.46. De Long and Summers's gap-based measure of performance, the ratio of the mean

deviation of GNP from potential, is 1.52 for the same sample periods. Clearly, both measures show about the same degree of stabilization over time.

Given that sensible volatility- and gap-based measures of macroeconomic performance show essentially the same amount of stabilization over time, it is clear that using gaps is not crucial to De Long and Summers's belief that we have tamed the business cycle. Rather, any apparent conflict between my work and that of De Long and Summers comes from how one interprets a stabilization ratio of 1.5. I tend to see it as small because I was used to seeing ratios of 2 or 3 when conventional data were used. Therefore, when my revised data showed a ratio below 1.5, I stressed that there was less stabilization than we once believed. On the other hand, De Long and Summers came up with a number of 1.5 and noticed that it was substantially bigger than one and therefore stressed that there had been some stabilization over time.

Both of these interpretations are defensible, but neither really deals with the question of how much stabilization a ratio of 1.5 actually represents. De Long and Summers seem to believe that any number bigger than one is important and represents a triumph for Keynesian stabilization policy and automatic stabilizers. But it is surely the case that at some point a small decline in the mean gap is not worth the potentially distortionary effects of procyclical taxation and misguided government expenditure. Furthermore, they provide no evidence that stabilization policy is what actually accounts for any stabilization that we observe. Policy could have stabilized the economy. But it is also possible that shocks to the economy could have been different in the two periods or that structural changes in the economy could have tended to improve macroeconomic performance.

In addition to questions about De Long and Summers's interpretation of the behavior of the output gap, one can also question their method for constructing a measure of the gap. Estimating potential GNP is very difficult and inevitably involves many choices. Throughout their analysis, nearly all of De Long and Summers's choices cause the postwar gap to be smaller than would result from an easily justifiable alternative choice. Thus, I would suggest that 1.5 should be viewed as an upper bound on a plausible ratio of average prewar to postwar gap, and that it would be easy to derive sensible gap-based measures that showed either no improvement or a worsening of macroeconomic performance over time.

Let me give a few examples of how their estimates of potential output are biased toward minimizing postwar gaps and maximizing prewar gaps. One of the criteria that they use in estimating potential is that actual output can never be above potential. They, however, exempt World War II from this requirement and assume that essentially all of the rise in output during the war was due to unusual wartime conditions. This assumption enables De Long and Summers to claim that the late 1940s and early 1950s were periods of excellent performance, equivalent in the absence of gaps to the late 1960s. However, it seems more appropriate to view at least some of the gains in output during World War II as evidence that potential was higher than the subsequent level of production in 1948 or 1950 might indicate.

A common argument for excluding World War II is that much of the increased production during the war was due to the entrance of women into the labor force in unprecedented numbers. That the labor force participation rate of women then decreases after the war is taken as a sign that such levels of production could not have been sustained. This common argument neglects the fact that after the war there was a concerted effort by policymakers and managers to get women out of the labor force—or at least out of "good jobs at good wages." Several authors have argued that women were forced out of jobs that they wished to keep because returning soldiers were deemed more worthy of employment.[1] If women wartime workers wished to keep working in highly productive factory jobs, then potential output should be seen as very high in the late 1940s and early 1950s, the gap should be viewed as very large, and policy should be viewed as responsible for large output losses. In this case gap-based measures of macroeconomic performance would show marked deterioration between the pre-Depression and post–World War II eras.

A related issue concerning wars and the measurement of potential output involves the treatment of World War I. While De Long and Summers are careful not to include World War II as a postwar peak, they do allow 1918 to be a peak. This naturally tends to make the gap in the early 1920s larger than it would have been if World War I had been eliminated from the sample. Furthermore, De Long and Summers accentuate the peak in 1918 by deflating output by the resident population

1. See Ruth Milkman, *Gender at Work* (University of Illinois Press, 1987), for an excellent description of the effects of demobilization on female wartime workers as well as references to other studies.

aged 16 and over, rather than the total population aged 16 and over, which would include persons serving in the armed services overseas. Given that GNP includes the output of the armed services (proxied by government spending), there is no justification for looking at GNP per resident person of working age rather than per person of working age. If this alternative procedure were followed, both the percentage change and the gap in 1920 and 1921 would be substantially smaller than De Long and Summers find. This would tend to make the prewar era look more like the postwar era than their numbers indicate.

Before leaving the empirical issues involved in measuring the output gap, it is important to mention De Long and Summers's use of European data to validate the decline in the U.S. gap over time. In contrast to Steven Sheffrin, who finds, using a volatility-based measure, that European countries have not stabilized over time, De Long and Summers conclude that various industrial democracies have shown an improvement in macroeconomic performance between the prewar and postwar eras.[2] However, this finding does not come from De Long and Summers's use of gaps, but from the fact that they stop their postwar sample in 1979. This exclusion of the 1980s is unjustifiable and is in no way similar to the exclusion of the 1930s. The Great Depression can rightly be left out of comparisons between the prewar and postwar eras because we do not know to which era it actually belongs. If one is using the prewar and postwar eras to proxy for the periods before and after activist government policy, a case can be made that the Depression belongs in the postwar era because both monetary policy and fiscal policy first started being used on a substantial scale during and after World War I. The same kind of argument cannot be made for the current high unemployment in Great Britain and other European countries. Surely these countries have not forgotten the policies that De Long and Summers think account for their fine performance before 1979.

My remaining comments concern the other types of evidence that De Long and Summers offer in support of their view that macroeconomic performance has improved markedly. First, they argue that a decline in the skewness of the unemployment rate series is evidence that postwar policy has filled in troughs without shaving off peaks. The empirical

2. Steven M. Sheffrin, "Have Economic Fluctuations Been Dampened?" *Journal of Monetary Economics*, vol. 21 (January 1988), pp. 73–83.

evidence of such a decline in skewness, however, is weak. The apparent skewness of the Lebergott and Romer prewar series shown in figure 2 of the paper is due entirely to the high unemployment rates of the 1890s. This is important because both Lebergott and Romer stress that the unemployment estimates for the 1890s are based on much sketchier information than are later estimates. In this regard, the fact that the Weir unemployment series does not show significantly more skewness than the modern series is telling because one of Weir's main contributions is to improve the unemployment series for the 1890s.[3] Given the behavior of the Weir series, one would have to conclude that there is no evidence of a decrease in the skewness of unemployment over time.

The final type of evidence about improvement in macroeconomic behavior that De Long and Summers consider is the possible decline in the persistence of fluctuations in real GNP over time. One of the main pieces of evidence that they invoke to suggest that prewar fluctuations were transitory in a way that postwar fluctuations are not is that we recovered from the Great Depression. Leaving aside the question of whether the Great Depression can tell us anything about the prewar or postwar eras in general, I want to disagree with their view of the health of the economy in the late 1930s. De Long and Summers argue that the economy had recovered from the Depression quite substantially even before the United States entered World War II. This view seems difficult to reconcile with the fact that the BLS estimate of the unemployment rate is 17.2 percent in 1939 and 14.6 percent in 1940. Even Michael Darby's alternative estimates of the interwar unemployment rate are nearly 10 percent as late as 1940.[4] These figures suggest to me that without World War II, the economy quite possibly would never have made up the output losses of the 1930s. In this case the fact that we recovered from the Great Depression is merely evidence that large positive shocks sometimes follow large negative shocks, even if the underlying process is white noise.

More generally, I am skeptical that their evidence on persistence provides important evidence about the effectiveness of postwar stabili-

3. David R. Weir, "Unemployment Volatility, 1890–1984: A Sensitivity Analysis" (Yale University, 1986).

4. Michael R. Darby, "Three-and-a-Half Million U.S. Employees Have Been Mislaid: Or, an Explanation of Unemployment, 1934–1941," *Journal of Political Economy*, vol. 84 (February 1976), pp. 1–16.

zation policy. First, it is not clear that successful stabilization policy should be expected to make output movements more persistent. Successful policy would reduce the importance of transitory cyclical movements. But as De Long and Summers have argued elsewhere, because policy should be best able to prevent *predictable* cyclical movements, successful policy would be likely to make the cyclical movements that remain less persistent.[5] The overall effect of these two influences on persistence is ambiguous. And, as with stabilization itself, changes in factors other than policy, such as institutions and the size and form of technology shocks, could also affect persistence.

Furthermore, given that De Long and Summers and I already agree that there has been some stabilization, any evidence on persistence is essentially impossible to interpret. The central issue is how much stabilization there has been. Does the change in persistence that De Long and Summers believe has occurred suggest a large or a small amount of stabilization? De Long and Summers provide no way of addressing this issue.

To conclude, I would have to say that despite the promising topic, the paper by De Long and Summers is ultimately unsatisfying. Using gaps does not materially alter the conclusions drawn from other, less subjective measures of macroeconomic performance, and the evidence on skewness and persistence is subject to alternative interpretations. More important, De Long and Summers provide no way of evaluating whether any improvement in macroeconomic performance that may have occurred is large or small and whether any of the change could have been due to policy. They seem content to say that since policy could have caused it, we should conclude that policy was effective. This argument, I'm afraid, will never convince anyone who does not already believe.

General Discussion

Laurence Ball supported the empirical proposition that demand shocks have asymmetric effects on output and that stabilization can therefore affect average output, but found the theoretical explanations

5. J. Bradford De Long and Lawrence H. Summers, "The Changing Cyclical Variability of Economic Activity in the United States," in Robert J. Gordon, ed., *The American Business Cycle: Continuity and Change* (University of Chicago Press, 1986), p. 697.

offered by De Long and Summers for such asymmetries inadequate. Ball was particularly critical of the models of price rigidity that De Long and Summers offered as explanations of the asymmetric effects of demand shocks. He noted that fixed-price models with rationing are inadequate theoretical models of price rigidity because the rigidity is added exogenously. More rigorous models of price rigidity, featuring price setters with market power, do not imply that price setters will be more likely to adjust prices up rather than down. Similarly, Ball noted that there is no theoretical model explaining why prices should be adjusted more quickly upward than downward. Theories in which demand shocks cause credit collapses display asymmetric output effects because, between collapses, the economy runs at potential. However, Ball observed that credit collapses do not explain most postwar recessions, though they may help explain the Great Depression.

David Romer suggested that stabilization policy does not need to raise mean output to have large welfare benefits. Although in the simple calculations of Lucas there is little gain from stabilization, if the benefits of stabilizing employment are considered, with an inelastic labor supply, the gains increase substantially. The necessary asymmetry enters through the utility function rather than through the aggregate supply function. Inelastic labor supply means that variation in average hours reduces utility; people do not get much pleasure out of their extra leisure during recessions, but suffer considerable disutility from the extra work required in a boom. Workers would require a large increase in real wages to compensate for working a few more hours, and they would not willingly choose to work fewer hours unless real wages plummet.

Edmund Phelps proposed a simple model explaining why workers voluntarily supply labor above the natural rate of employment. Employers set a real wage above the market-clearing level to minimize labor costs per efficiency unit of work. At the natural rate equilibrium, the real wage is high enough to generate involuntary unemployment, and the fear of unemployment prevents workers from shirking. A firm faced with a demand shock hires more workers either because it suspects the shock is affecting only its own demand or because the pay scale of the firm cannot be adjusted in the short run. If it is an aggregate demand shock, all firms increase employment, and the aggregate level of employment exceeds its natural rate. Phelps emphasized that in this model workers are happy to supply labor above the natural rate; it is the firms that are tricked into hiring them.

Robert Hall applauded De Long and Summers for abandoning the natural rate hypothesis, a theory whose acceptance, in his opinion, had been a victory of theory over fact. He noted that estimates of the natural rate of unemployment tend to track the actual unemployment rate, making the theory empirically empty. Phelps disagreed, arguing that the natural rate is relatively stable in the United States and, in any case, that shifts in the natural rate are not evidence against the theory. He noted that economists do not abandon the concept of money demand because estimates of money demand are unstable.

Hall went on to observe that real business cycle theory and some theories of multiple equilibriums abandon the natural rate. He recommended that the term "thick-market externalities" replace the term "demand externalities" as a way to identify new models of multiple equilibriums that predict that the economy will drift from one equilibrium to another. He found it misleading to suggest that in models with thick-market externalities, the economy is generally lodged at the high-output equilibrium, only occasionally slipping to lower-level equilibriums. However, the models are consistent with the proposition that successful monetary policy might ensure that the economy lodges at the high-level equilibrium.

Hall explained that the central implication of models with thick-market externalities is that economic activity will be bunched, most obviously during weekdays, daylight hours, and the Christmas season. The central question is whether business cycles are actually an optimal bunching of economic activity. If so, this would undermine the basic proposition of De Long and Summers that the economy ought to be stabilized at the high-level equilibrium. Hall noted that wars represent an extreme example of the bunching of economic activity. He agreed with Christina Romer that wars also represent a challenge to the output gap methodology. Even if the wages of draftees are subtracted from GNP, De Long and Summers would still show a large negative gap during World War II. Hall concluded that the economics of World War II deserves as much attention as the economics of the Depression.

George von Furstenberg suggested that the main effect of government policy was often to reduce the mean level of output through misguided microeconomic policies. He pointed to the requirement in many European countries that firms provide extensive insurance for their employees. A European firm that hires an extra employee must not only pay a

wage but also assume an additional contingent liability that becomes due if the firm ever wishes to lay off the employee. Von Furstenberg noted that these indirect labor costs, which may be nearly as large as direct wage costs, contribute to the high unemployment rate in Europe by inhibiting firms from hiring additional employees unless they perceive a nearly permanent increase in demand.

Matthew Shapiro criticized the presumption that good government policy could have kept the level of output at the peak-to-peak measures of potential output drawn by De Long and Summers. Shapiro reasoned the measure of potential output has to be related to the technological capacity of the economy and proposed a more structural analysis that looks at the behavior of capacity and labor input as well as output. He noted further that the finding that output has become more persistent in the postwar period does not necessarily imply that government policy has become more effective. The finding is also consistent with Christina Romer's view that transitory measurement error has been reduced.

Robert Gordon was surprised that De Long and Summers never mention prices even though Summers has been highly critical of real business cycle models that ignore prices. Gordon contended that the empirical validity of the natural rate hypothesis comes from its ability to explain trends in aggregate prices, such as the increasing rates of inflation over the 1960s. He called attention to new findings on prices before the Depression that support the argument that post–World War II government policy has increased average output. Previous researchers had looked at deflators created from crude and intermediate materials prices and concluded that pre–World War I prices were quite flexible. However, data on prices paid by consumers before World War I, collected by Al Rees in the 1950s, suggest that prices were actually quite sticky. That inflation did not accelerate in the prosperous period from 1901 to 1906, in the period before World War I, or in the relatively prosperous years of the mid-1920s suggests that the economy never reached potential output in any of those early periods.

Gordon noted that measures of potential output can be derived from Okun's Law using a measure of the unemployment rate. Such measures of potential show substantial changes in growth rates between benchmark years. Therefore, Gordon was not persuaded by the evidence that the unemployment rate is correlated more strongly with De Long and Summers's output gap measure than with measures of deviations from

trend. It is not clear whether the higher correlation is due to the superiority of a gap methodology to a cycle methodology or simply to the breaks in the growth rate of the peak-to-peak measure of potential used to construct the gap.

Martin Baily observed that the Depression is really the primary evidence in favor of De Long and Summers's criticism of natural rate theory. Taken literally, equation 2, which is derived from the natural rate hypothesis, implies that the economy should have emerged from the Depression with a legacy of deflation and hence a chance to grow for years to come without fear of inflation. In fact, the economy did not seem to gain any such benefit from the Depression. Baily reasoned that stabilization policy should be designed to avoid persistent downturns such as the Depression. He interpreted the current high unemployment rates in Europe as a sign that European governments have abandoned even this modest version of stabilization policy.

Gordon criticized Christina Romer's suggestion that monetary or fiscal policy could be interpreted as activist during the Depression. He noted that in the late 1930s the idea of "pushing on a string" was invented to argue that monetary policy was ineffective, and cited E. Cary Brown's finding that the full-employment government surplus, including all levels of government, was contractionary during the whole Roosevelt administration up through 1940. Romer responded that even though government policy may have been used perversely during the Depression, the tools of government policy, such as open market operations and the ability to increase government spending, were available.