Federal Reserve Bank of St. Louis

REVIEW

JULY/AUGUST 2005 VOLUME 87, NUMBER 4

Productivity, Labor, and the Business Cycle

Proceedings of the Twenty-Ninth Annual Economic Policy Conference of the Federal Reserve Bank of St. Louis

Productivity, Labor, and the Business Cycle

Review is published six times per year by the Research Division of the Federal Reserve Bank of St. Louis and may be accessed through our web site: **research.stlouisfed.org/publications/review**. All nonproprietary and nonconfidential data and programs for the articles written by Federal Reserve Bank of St. Louis staff and published in *Review* also are available to our readers on this web site. These data and programs are also available through Inter-university Consortium for Political and Social Research (ICPSR) via their FTP site: www.icpsr.umich.edu/pra/index.html. Or contact the ICPSR at P.O. Box 1248, Ann Arbor, MI 48106-1248; 734-647-5000; **netmail@icpsr.umich.edu**.

Single-copy subscriptions are available free of charge. Send requests to: Federal Reserve Bank of St. Louis, Public Affairs Department, P.O. Box 442, St. Louis, MO 63166-0442, or call (314) 444-8809

General data can be obtained through FRED (Federal Reserve Economic Data), a database providing U.S. economic and financial data and regional data for the Eighth Federal Reserve District. You may access FRED through our web site: **research.stlouisfed.org/fred**.

ISSN 0014-9187

Contributing Authors

Fernando Alvarez
Department of Economics
University of Chicago
Rosenwald Hall, Office 318-G
1101 East 58th Street
Chicago, IL 60637
f-alvarez1@uchicago.edu

Ricardo J. Caballero
Department of Economics
Massachusetts Institute of Technology
E52-252a
50 Memorial Dr.
Cambridge, MA 02142-1347
caball@mit.edu

Jordi Galí
CREi
Ramon Trias Fargas, 25-27
08005 Barcelona
Spain
jordi.gali@upf.edu

James D. Hamilton
Department of Economics
University of California, San Diego
La Jolla, CA 92093-0508
jhamilton@ucsd.edu

Robert G. King
Department of Economics
Boston University
270 Bay State Road
Boston, MA 02215
rking@bu.edu

Kathryn Koenders
Department of Economics
Arizona State University
Tempe, AZ 85287
kathryn.koenders@asu.edu

Mau-Ting Lin
Department of Economics
National University of Singapore
Faculty of Arts and Social Sciences
Bock AS2, Level 6
1 Arts Link, nk
Singapore
ecslmt@nus.edu.sg

Ellen R. McGrattan
Research Department
Federal Reserve Bank of Minneapolis
90 Hennepin Avenue
Minneapolis, MN 55480-0291
erm@ellen.mpls.frb.fed.us

Edward C. Prescott
Department of Economics
W. P. Carey School of Business
Arizona State University
Tempe, AZ 85287-3806
edward.prescott@asu.edu

Richard Rogerson
Department of Economics
Arizona State University
Tempe, AZ 85287
richard.rogerson@asu.edu

Julio J. Rotemberg
Harvard Business School
Morgan 297
Boston, MA 02163
jrotemberg@hbs.edu

Robert Shimer
Department of Economics
University of Chicago
1126 East 59th Street
Chicago, IL 60637
shimer@uchicago.edu

Christopher A. Sims
Department of Economics
Princeton University
104 Fisher Hall
Princeton, NJ 08544-1021
sims@princeton.edu

Mark W. Watson
Department of Economics
Princeton University
413 Robertson Hall
Princeton, NJ 08544-1013
mwatson@princeton.edu

Randall Wright
Department of Economics
University of Pennsylvania
3718 Locust Walk
Philadelphia, PA 19104-6297
rwright@econ.upenn.edu

President's Message

William Poole

The driving force behind economic growth is productivity, a product of embodied technological progress and capital deepening. A number of other factors can affect productivity and, thus, the trajectory of the economy, including demographics and labor force participation. Structural change in the labor market, spurred by innovations in communication or changes in regulations, can also dramatically alter the landscape. Understanding these factors is critical to economic policymaking.

As is well known, U.S. productivity growth has been far from constant. Prior to World War II, productivity growth averaged about 2 percent per year. From the end of the war to 1973, productivity growth averaged about 3 percent per year. Then we suffered the *Great Productivity Slowdown*, during which productivity growth fell precipitously to an average of 1.5 percent per year from 1973 to 1995. A revival started in 1995; from that year to date, productivity growth has averaged 2.9 percent per year. Remarkably, the recent recession and slow recovery has not dampened productivity growth, which has averaged 4.0 percent since the first quarter of 2001.

Compound interest is the eighth wonder of the world: If labor productivity continues to grow at 3 percent per year, real per capita gross domestic product (GDP) in the United States will increase from $35,700 in 2003 to $48,000 by 2013. Conversely, at the slower 1.5 percent growth rate, real per capita GDP will rise to only $41,400, nearly $7,000 per person less.

Fluctuations in productivity growth give rise to a number of important questions, some of which we hope to address in this conference. Specifically, what factors lead to shifts in productivity growth and how these are different from the forces that create cyclical fluctuations. We will investigate whether these factors we know to be key determinants of long-run growth can also be the primary influences for cyclical fluctuations.

Many economists have argued that the shift in productivity growth in 1995 occurred because of a recent increase in investment in, and implementation of, information technology (IT). The share of GDP tied to investment in the IT sector is now more than 50 times what it was in 1975. Increasing productivity in the IT sector has spurred sharp declines in the prices of computers, helping to drive investment in the late 1990s. Aside from the direct impact this IT capital investment has on enhancing labor productivity, the rapid rate of investment leads to a rise in productivity growth through capital-deepening.

But increased investment in IT alone cannot fully explain all of the business cycle phenomena we have experienced throughout the postwar period. As we will see over the course of this conference, a myriad of other factors can influence the behavior of both long-run growth and cyclical fluctuations. For example, the structure of the labor market and the rate of organizational restructuring during booms and recessions can affect the cyclicality of key business cycle indicators and perhaps help explain the occurrence of the recent "jobless" recoveries. We will learn whether these jobless recoveries represent, in fact, aberrant business cycle behavior or whether we can draw inference from economic history to explain their

William Poole is the president of the Federal Reserve Bank of St. Louis.

Federal Reserve Bank of St. Louis *Review*, July/August 2005, *87*(4), pp. 427-28.

existence. Finally, we will investigate how the reaction of monetary policy in the presence of such a changing environment influences real outcomes and that, perhaps, the familiar Taylor-type monetary reaction functions we have become used to as policymakers might lead to results different from what we expect.

So, I welcome you to the St. Louis Fed's annual research conference and will now get out of the way so the fun can begin.

Editor's Introduction

Michael T. Owyang

A policymaker's decision process involves a sequence of evaluations: What is the true state of the economy, why is the economy in this state, and how would policy affect the economy? This "what, why, how" sequence of evaluations requires a complete toolbox from the economist. On October 21 and 22, 2004, the Federal Reserve Bank of St. Louis hosted its Twenty-Ninth Annual Economic Policy Conference under the broad umbrella topic "Productivity, Labor, and the Business Cycle." At the conference, six diverse papers were presented that focused on different aspects of the business cycle. Indeed, these papers exhibited the breadth of the economist's toolbox, ranging from a statistical representation of the business cycle, to an econometric analysis of its driving forces, to a measurement of key indicators, to theoretical models of its causes and effects.

WHAT IS THE BUSINESS CYCLE?

The notion that the economy inhabits distinct phases traces back to Burns and Mitchell (1946). Recent developments in econometrics have led economists to question whether these phases can be characterized by nonlinear statistical models. In his conference paper, James Hamilton asks whether nonlinear models, specifically Markov-switching models, provide evidence of a "real" business cycle. The advantage of Markov-switching models is they can allow for changes in dynamics across states without imposing strict periodicity.

In a previous paper, Hamilton (1989) considered a two-state Markov-switching model of U.S. gross domestic product (GDP) and found that the timing of switches roughly coincided with turning points determined by nonstatistical methods (e.g., the NBER Business Cycle Dating Committee). In his paper for this conference, Hamilton posits a three-state Markov-switching intercept model with non-Gaussian innovations for unemployment. This third state is revealed to be one of exceptionally high unemployment, which occurs infrequently and never directly following the expansionary state.

Hamilton also considers nonlinearities in interest rates. His sample includes the latter half of the nineteenth and the beginning of the twentieth centuries and ends before the establishment of the Federal Reserve System. Here, Hamilton employs a model with two-state Markov switching in the variance of the innovation: He shows that, although not every recession seems correlated with an increase in interest rate volatility, several pre-Federal Reserve recessions were nearly coincident with shifts in the variability of interest rates.

In his comments on the Hamilton paper, Mark Watson concentrates on two econometric issues: (i) Are nonlinearities necessary to model business cycles? and (ii) Do nonlinearities aid in forecasting business cycle variables? Watson addresses the first question by considering four series: U.S. GDP and three artificially constructed series simulated from a calibrated AR(2) process. These series illustrate results he attributes to Slutsky (1937)—namely, that the realizations from a linear model can behave in a manner that generates cycles and *appears* nonlinear. Watson's

Michael T. Owyang is a senior economist at the Federal Reserve Bank of St. Louis.

Federal Reserve Bank of St. Louis *Review*, July/August 2005, *87*(4), pp. 429-33.

conclusion is that nonlinearities are not necessary a priori to have business cycles but in fact govern our characterization of the data.

Watson addresses the second question by investigating whether the filtered probabilities (i.e., the turning points) estimated from Hamilton's three-state Markov-switching model forecast other business cycle variables. Concentrating on the filtered probabilities for unemployment, Watson finds some evidence that Hamilton's estimated turning points forecast business cycle variables such as industrial production and personal income.

TECHNOLOGY SHOCKS AND THE BUSINESS CYCLE

In a previous paper, Jordi Galí (1999) called into question an assumption of real business cycle theory: that technology shocks are a driving force in cyclical fluctuations. In that paper, Galí estimated a structural vector autoregression (VAR) with the identifying assumption that technology was the only shock that could affect labor productivity in the long run. Among other findings, he discovered that the short-run response of hours to a positive technology shock was negative. However, some recent studies have called into question Galí's specification of hours as an I(1) series in his VAR.

In his paper in this volume, Galí attempts to reconcile the specification of hours as nonstationary by proposing a benchmark balanced-growth model. He shows that the nature of the hours series depends on the (non)stationarity of the consumption share. Moreover, he argues that low-frequency fluctuations can confound estimation of the short-run responses to shocks. He presents evidence from the G7 countries to support his claim that the labor input can/should be modeled as nonstationary. Finally, he reestimates the identified VAR from Galí (1999) for the G7 countries, substituting hours in differences for employment in differences in each country's VAR.

In his discussion of the Galí paper, Chris Sims contrasts the conclusions drawn by the VAR literature with those arising from more-complicated structural models using Bayesian methods. Sims cautions that small models such as Galí's VAR may not be rich enough to account for low-frequency dynamics. Sims's concerns about the methodology are summarized by two questions: (1) Are long-run restrictions of the kind employed by Galí sufficient to achieve identification? and (ii) How robust are Galí's results to alternative assumptions about the unit-root behavior of hours? Sims concludes that long-run restrictions provide only a weak identification. Moreover, he shows that adjusting the strength of the priors of unit roots in both labor productivity and hours has significant effects on the posterior distributions for the impulse responses and, thus, also on the conclusions that Galí draws.

UNEMPLOYMENT AND THE BUSINESS CYCLE

One feature that defines the business cycle is the acyclicality of unemployment. The unemployment rate, however, is jointly determined by hires and separations—the flow of workers both between jobs and between employment and unemployment. In his paper in this volume, Robert Shimer constructs a model of employment dynamics with on-the-job search in an effort to determine the cyclicality of job flows. First, he proposes a method to estimate the rates of job finding and separation. He constructs simple accounting rules that can be used on data taken from the Current Population Survey to compute these rates.

Second, Shimer proposes a model with which he can estimate the rate of job-to-job transitions. He argues that workers move from job to job only if the new job is of higher quality (i.e., if the sum of the pecuniary and nonpecuniary benefits is greater). From this theory, Shimer computes theoretical job-to-job transitions rates, which he compares with three measures generated by methods in the current literature and finds they are roughly consistent. Finally, Shimer concludes that it is the rate at which workers, employed and unemployed, find jobs that dictates fluctuations in unemployment.

In his discussion of the Shimer paper, Randy

Wright concentrates on issues of measurement. He acknowledges Shimer's attempts to correct for data problems but points out some instances where data errors may lead to mismeasurement of hiring and separation rates. First, what is the real effect of unmeasured heterogeneity? Wright argues that accounting for race, sex, and other demographic characteristics may be insufficient to correct for all forms of heterogeneity. In particular, he cites laziness as an unobserved quality that would affect productivity but would go unobserved by the econometrician. Second, Wright asks about the potential bias that might arise if erstwhile employees who exit the labor force are not accounted for. Finally, he asks whether time aggregation issues cause biases in the computation of job-to-job transitions. Some of the apparent job-to-unemployment-to-job transitions may actually be job-to-job movements in which the worker intentionally takes time off.

MONETARY POLICY AND THE BUSINESS CYCLE

The role of monetary policy in either exacerbating or damping cyclical fluctuations is explored in the conference paper by Robert King and Mau-Ting Lin. They investigate the responses of prices and output to shocks to government spending and productivity under alternative interest rate rules. The three policy regimes they consider involve rules in which interest rates respond to (i) inflation only, (ii) inflation and the output gap with weights taken from Taylor (1993), and (iii) inflation, the output gap, and past interest rates with weights taken from estimates constructed by Orphanides and Wieland (1998).

The model employed by King and Lin is similar to the representative agent model of King and Wolman (1996), which appeared in an earlier issue of this *Review*. Similar to that model, King and Lin include capital adjustment costs and monopolistically competitive firms. However, King and Lin do not assume a shopping time constraint and, thus, have no explicit assumption about the demand for money. Their model includes a monetary policymaker who sets interest rates according to a predetermined rule. It is the nature and effect of this rule that is of interest in this paper. King and Lin show that, in their model, the rule that responds only to inflation amplifies cyclical fluctuations. On the other hand, the two rules that incorporate both inflation and the output gap reduce the magnitudes of cycles. These results are then examined in the context of the previous literature on monetarist policy prescriptions.

In his discussion of the King and Lin paper, Julio Rotemberg cautions that the current framework may have subtle differences from the monetarist ideals. In particular, the monetarists wish to achieve output stability, yet the King and Lin framework will yield output fluctuations upon any innovation to the IS curve (e.g., government purchases). Moreover, Rotemberg questions whether fluctuations in the King and Lin model arise from innovations to the real variables (i.e., technology and government purchases) or whether they are caused by shocks to monetary policy. He argues that government purchase shocks, because they affect markups, may also feed through the innovation in the policy equation. Rotemberg then considers the role of technology shocks in the King and Lin model. He posits that stabilizing output around trend is desirable only if the underlying trend can be ascertained by the central bank.

MEASURING THE BUSINESS CYCLE

Labor productivity, typically computed as GDP per hour worked, is a commonly cited measure of the welfare of the economy. In their paper in this volume, Ellen McGrattan and Edward Prescott argue that this measure of productivity can be misleading. In particular, they contend that GDP per hour worked may not fully represent the booming productivity growth in the late 1990s and that a true measure of economic productivity for that period should be substantially higher. They contend that this differential arises from an accounting problem in which some investment, which they term intangible investment, is neglected in the measure of GDP.

McGrattan and Prescott propose an alternative measure of economic productivity that accounts for intangible investment. By constructing a representative agent model that explicitly accounts for corporate profits, they estimate the average value of intangible capital during the 1990s and, perhaps more importantly, show that intangible capital rose substantially during the late 1990s. By taking seriously this unmeasured economic productivity, the authors characterize the late 1990s as a period of high prosperity, beyond the level held in the conventional view.

In his discussion of the McGrattan and Prescott paper, Ricardo Caballero argues that the introduction of short-run frictions (e.g., investment adjustment costs and labor mobility frictions) can bias computation of intangible capital. In particular, Caballero shows that McGrattan and Prescott's correction for intangible capital may overestimate the acceleration of intangible investment in the late 1990s. He posits an alternative accounting adjustment that characterizes intangible investment in the 1990s as potentially intertemporally substituted away from the mid-1990s to the late 1990s. In other words, the rise in intangible investment in the late 1990s simply compensates for a decline in the mid-1990s. Moreover, this adjustment may mitigate the increase in economic productivity advocated by McGrattan and Prescott.

ORGANIZATIONAL BEHAVIOR AND THE BUSINESS CYCLE

Recoveries from postwar recessions have generally been characterized by strong employment growth at a one-quarter lag from the turnaround in GDP. In the two most recent recessions (1990 and 2001), however, the decline in employment has been more persistent (i.e., employment has taken substantially longer to recover to prerecession levels). In their paper in this volume, Kathryn Koenders and Richard Rogerson argue that changes in organizational restructuring may have contributed to the so-called jobless recovery of the past two recessions. They construct a model in which a manager chooses between production and reorganization to reduce organizational inefficiency. They show that this reorganization occurs predominantly in the recession and recovery periods of the business cycle, a time in which the opportunity cost of reorganization is relatively low.

Next, Koenders and Rogerson consider the evidence for jobless recoveries over the past eight postwar recessions. Consistent with their model, they find that productivity drops during the recession, suggesting a period of reorganization. Then, by first detrending employment, they show that the jobless recovery may not be unique to the past two recessions. In fact, a similar persistent decline in employment followed the 1970 recession.

In his discussion of the Koenders and Rogerson paper, Fernando Alvarez extends the model to general equilibrium. His goal is to determine whether the results from the partial equilibrium model are indeed consistent with the reduced-form planning problem that would obtain in a general equilibrium framework. He analyzes the manager's response to an i.i.d. demand shock and shows that reorganization is countercyclical, a result consistent with the partial equilibrium framework outlined in the Koenders and Rogerson paper.

Finally, I would like to thank the authors and discussants for their papers and participation in the conference. In addition, I would like to express my appreciation to all the conference participants. I would also like to thank the Bank's research staff, especially Kristie Engemann, Heidi Beyer-Powe, and Beverly Benham, for their assistance in organizing the conference.

REFERENCES

Burns, Arthur F. and Mitchell, Wesley C. *Measuring Business Cycles*. New York: National Bureau of Economic Research, 1946.

Galí, Jordi. "Technology, Employment, and the Business Cycle: Do Technology Shocks Explain Aggregate Fluctuations?" *American Economic Review*, 1999, *89*(1), pp. 249-71.

Hamilton, James D. "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle." *Econometrica*, March 1989, *57*(2). pp. 357-84.

King, Robert G. and Wolman, Alexander L. "Inflation Targeting in a St. Louis Model of the 21st Century." Federal Reserve Bank of St. Louis *Review*, May/June 1996, *78*(3), pp. 83-107.

Orphanides, Athanasios and Wieland, Volker. "Price Stability and Monetary Policy Effectiveness when Nominal Interest Rates Are Bounded at Zero." Finance and Economics Discussion Series, 1998-35; Board of Governors of the Federal Reserve System, 1998; www.federalreserve.gov/pubs/feds/1998/199835/199835abs.html.

Slutsky, Eugen. "The Summation of Random Causes as the Source of Cyclic Processes." *Econometrica*, April 1937, *5*(2), pp. 105-46.

Taylor, John B. "Discretion Versus Policy Rules in Practice." *Carnegie-Rochester Conference Series on Public Policy*, 1993, *39*, pp. 195-214.

What's Real About the Business Cycle?

James D. Hamilton

This paper argues that a linear statistical model with homoskedastic errors cannot capture the nineteenth-century notion of a recurring cyclical pattern in key economic aggregates. A simple nonlinear alternative is proposed and used to illustrate that the dynamic behavior of unemployment seems to change over the business cycle, with the unemployment rate rising more quickly than it falls. Furthermore, many but not all economic downturns are also accompanied by a dramatic change in the dynamic behavior of short-term interest rates. It is suggested that these nonlinearities are most naturally interpreted as resulting from short-run failures in the employment and credit markets and that understanding these short-run failures is the key to understanding the nature of the business cycle.

Federal Reserve Bank of St. Louis *Review*, July/August 2005, *87*(4), pp. 435-52.

WHAT IS THE BUSINESS CYCLE?

The term "cycle" is used to describe a process that moves sequentially between a series of clearly identifiable phases in a recurrent or periodic fashion. Economists of the nineteenth and early twentieth centuries were persuaded that they saw such a pattern exhibited in the overall level of economic activity and enthusiastically sought to characterize the observed regularities of what came to be known as the "business cycle." The most systematic and still-enduring summaries of what seems to happen during the respective phases were provided by Mitchell (1927, 1951) and Burns and Mitchell (1946).

The expression "business cycle theory" remains in common usage today, even though, in most of the modern models that wear the label, there in fact is no business cycle in the sense just described. These are models of economic fluctuations, to be sure, but they do not exhibit clearly articulated phases through which the economy could be said to pass in a recurrent pattern.

In part, this shift in the profession's conception of what needs to be explained about business fluctuations reflects a desire to integrate the determinants of long-run economic growth and the causes of short-run economic downturns within a single unified theory of aggregate economic performance. Since improvements in overall productivity are widely acknowledged to be one of the key factors driving long-run growth, and since such improvements cannot reasonably be expected to occur at a constant rate over time, it is natural to explore the possibility that variation over time in the rate of technological progress could be a primary cause of variation over time in the level of economic activity. Brock and Mirman (1972) were the first to incorporate stochastic variation in the rate of technical progress into a neoclassical growth model, though they clearly intended this as a model of long-run growth rather than a realistic description of short-run fluctuations. Kydland and Prescott (1982) later took the much bolder step of proposing that this class of models might explain variations in economic activity at all fre-

James D. Hamilton is a professor of economics at the University of California, San Diego. This research was supported by the National Science Foundation under grant No. SES-0215754.

quencies, in what has come to be known as "real business cycle models."

Although unifying growth and business cycle theory holds tremendous aesthetic appeal, this particular solution is not without its detractors. Indeed, the reasons that Irving Fisher gave in 1932 for rejecting such an approach have in the opinion of many yet to receive a satisfying response from modern real business cycle theorists:

> [I]n times of depression, is the soil less fertile? Not at all. Does it lack rain? Not at all. Are the mines exhausted? No, they can perhaps pour out even more than the old volume of ore, if anyone will buy. Are the factories, then, lamed in some way—down at the heel? No; machinery and invention may be at the very peak. (Fisher, 1932, p. 5)

Continuing along the lines of Fisher's reasoning, the size of the population places an obvious physical limit on how much a given nation can produce and is certainly a key reason that aggregate output increases over time. But just as surely, a decrease in population is not the cause of the decrease in employment that we observe in times when the unemployment rate is shooting up dramatically. There is in this respect an obvious inherent asymmetry in fluctuations in the number of workers employed—the measure must go up for different reasons than it goes down. A parallel argument can be made in terms of the capital stock, another key factor determining long-run growth, which again places an upper limit on how much a country can produce. Yet in times when we see all measures of capacity utilization falling, the natural inference is that some forces other than the quantity or quality of available manufacturing facilities account for the drop in aggregate output.

If we agree that these three factors—technology, labor force, and the capital stock—are the three main determinants of long-run economic growth, we might greet with considerable skepticism the suggestion that the same three factors are in a parallel way responsible for producing the drop in real GDP that we observe during a business downturn.

The purpose of this paper is to explore whether the nineteenth-century economists were on to something that their modern descendants may have forgotten. Is there really a business cycle, or is the expression an unfortunate linguistic vestige of a less-informed era? I will argue that indeed there is a recurring pattern in the level of economic activity that needs to be explained, but that a statistical characterization of this pattern requires a nonlinear dynamic representation and calls for an asymmetric interpretation of the forces that cause employment to rise and fall. I further observe that one element of this pattern has often been a related cyclical behavior of interest rates.

To the question, "Is the business cycle real?" these findings suggest that, yes, the business cycle is real in the sense that it is a feature of the data that needs to be explained. In the other meaning of the term "real," however—the sense from which springs the label "real business cycle," namely, a cycle unrelated to monetary developments—the evidence adduced here for the importance of comovements between financial and real variables suggests that the cycle is not "real" at all or, at the least, not completely divorced from monetary developments.

THE BEHAVIOR OF UNEMPLOYMENT

Figure 1 plots the monthly unemployment rate in the United States from 1948:01 to 2004:03.[1] I would suggest that someone looking at such a graph for the first time would indeed be inclined to identify a repeated sequence of ups and downs, with each of the obvious sharp upswings in the unemployment rate occurring during periods that the National Bureau of Economic Research (NBER) has classified as economic recessions (indicated by shaded regions on the graph).

Although one's eye is sympathetic to the claim that these data display a recurrent pattern, it does not appear to be cyclical in the sense of exhibiting strict periodicity. For example, the two consecutive unemployment peaks in 1958:07 and 1961:05 are separated by less than three years, whereas

[1] This is the seasonally adjusted civilian unemployment rate from the Bureau of Labor Statistics; http://stats.bls.gov.

Figure 1

U.S. Monthly Civilian Unemployment Rate and U.S. Recessions, 1948:01–2004:03

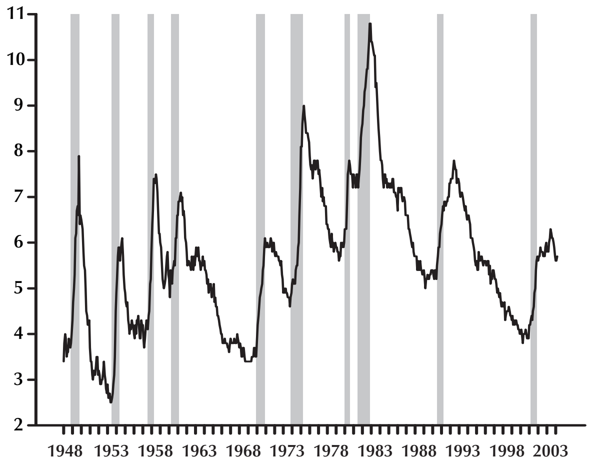
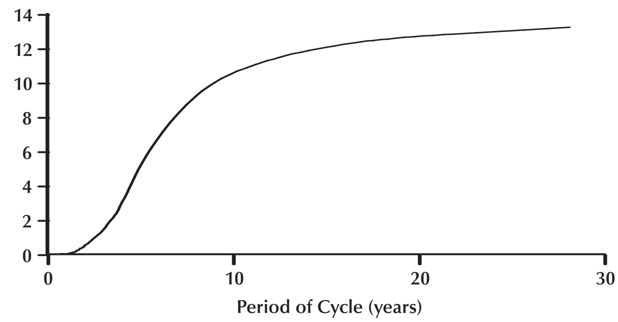


Figure 2

Estimated Spectrum of U.S. Monthly Civilian Unemployment Rate, 1948:01–2004:03



NOTE: Plotted as a function of the period of the cycle in years.

those of 1982:11 and 1992:06 are separated by a decade. More formally, one can look for any sort of periodic pattern by examining the spectrum of the unemployment rate, an estimate of which is plotted in Figure 2 as a function of the period of the cycle.[2] If one tries to decompose the unemployment series in Figure 1 into a series of strictly periodic cycles, by far the most important of these are those with the longest period, as opposed to something regularly repeating every 3 to 5 years.

Let y_t denote the unemployment rate. Consider an AR(2) representation of these data with Student t innovations, obtained by maximizing the log likelihood function

$$\mathfrak{L}(\theta) = \sum_{t=3}^{T} \ell_t(\theta)$$

$$\ell_t(\theta) = k - \left[(\nu+1)/2\right]\log\left[1 + \frac{u_t^2}{\nu\sigma^2}\right] \tag{1}$$

$$u_t = y_t - c - \phi_1 y_{t-1} - \phi_2 y_{t-2} \tag{2}$$

$$\begin{aligned} k = {} & \log\left\{\Gamma\left[(\nu+1)/2\right]\right\} - \log\left\{\Gamma\left[\nu/2\right]\right\} \\ & -(1/2)\log\left(\sigma^2\nu\pi\right) \end{aligned} \tag{3}$$

with respect to $\theta = (c, \phi_1, \phi_2, \sigma^2, \nu)'$ subject to the constraints[3] that $\sigma^2 > 0$ and $\nu > 0$. These maximum likelihood estimates (MLEs) (with asymptotic standard errors in parentheses) imply that the unemployment rate y_t for month t could be modeled as follows:

$$y_t + \underset{(0.028)}{0.060} + \underset{(0.037)}{1.117}\, y_{t-1} - \underset{(0.037)}{0.128}\, y_{t-2} + \underset{(0.007)}{0.158}\, v_t, \tag{4}$$

where v_t is distributed Student t with 4.42 degrees of freedom, with the standard error for the degrees-of-freedom parameter ν being estimated at 0.74. Using Student t innovations instead of Normal innovations increases the log likelihood by 52.04, a huge gain from estimating the single parameter ν (see Table 1).

As further evidence against a cycle with regular periodicity, it is interesting to note that the roots of the second-order difference equation in (4) are both positive and real, meaning that this system does not exhibit any oscillatory behavior in response to a shock to v_t.

---

[2] This was calculated by smoothing the sample periodogram with a Bartlett window (e.g., Hamilton, 1994, eq. [6.3.15]) with lag $q = 13$, as calculated using the RATS fft procedure with window (type = tent, width = 25). See the procedure hamp167.prg available at www.estima.com/procs_hamilton.shtml for details. The resulting estimate $\hat{s}_Y(\omega_j)$ for $\omega_j + 2\pi j/T$ is plotted in Figure 2 for given j as a function of T/j, which is the variable measured on the horizontal axis.

[3] See, for example, Hamilton (1994, Section 5.9) on numerical maximization subject to inequality constraints.

Table 1

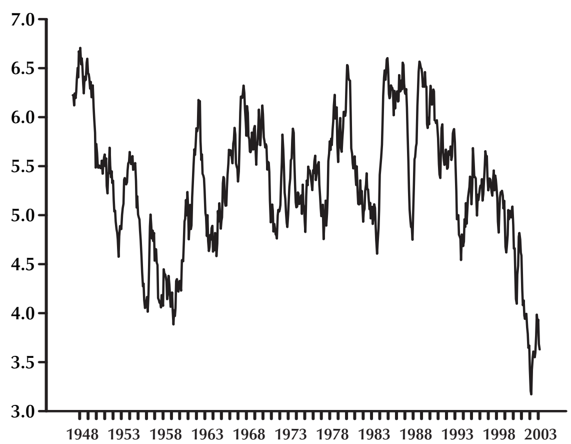
Comparison of Selected Models of Postwar Unemployment Rates

Model	No. of parameters	Log likelihood	Schwarz criterion
Gaussian AR(2)	4	75.59	62.57
Student *t* AR(2)	5	127.63	111.35
Student *t* AR(2) with MS intercept	11	174.58	138.77

NOTE: Schwarz criterion calculated as $\mathfrak{L} - (k/2)\log(T)$ for $\mathfrak{L}$ the log likelihood, k the number of parameters, and $T = 673$ the sample size.

Figure 3

Simulated Unemployment



NOTE: Simulated sample generated from equation (4).

Is the appearance of a repeated cycle in Figure 1 just a figment of our imagination, then? Another interesting exercise is to simulate a time-series realization from (4), which is displayed in Figure 3. These simulated data have the same mean, variance, and serial correlation as the real data in Figure 1, as of course they should. Even so, one has little of the sense of a recurrent cycle in these simulated data that seemed compelling in the actual data. If one were to label some of the episodes in this simulated data set as "recessions," where would they be? Indeed, expression (4) characterizes the true process from which these artificial data were simulated. What in terms of the qualities of this data-generating process would one characterize as a "business cycle?" There are good and bad values of the innovations v_t, and perhaps we could make up some rule for categorizing a relatively unlikely string of mostly negative innovations as a "recession." But any such rule would be completely arbitrary and tell us more about our imagination or quest for patterns and labels than about anything in the objective reality. There is nothing qualitatively different about a value of v_t that puts us within the arbitrary recession category and one that leaves us just short of it.

I would argue that this inability to define a business cycle as a fundamental attribute of the data-generating process (4) is in fact inherent in any time-series model that describes y_t as a linear function of its lagged values plus an i.i.d. innovation. Even if the linear difference equation did exhibit an oscillatory impulse-response function or imply more power in the spectrum at periods of 3 to 5 years, it seems to be some other feature of the data in Figure 1 that constitutes the "business cycle."

I would suggest instead that what we have in mind is that there is something in common between the rapid run-ups in unemployment that occurred in each of the postwar recessions, even though the length of time it takes for unemployment to spike up varies from episode to episode, and the timing separating such events is irregular. Indeed, the idea of looking for commonality across recessions whose elapsed calendar time is different for different episodes was precisely the methodology that Burns and Mitchell used to create their graphs summarizing typical business cycle patterns. Stock (1987, 1988) showed that such a way of thinking about data necessarily implies a nonlinear data-generating process.

Table 2

Maximum Likelihood Estimates of Three-State Markov-Switching Model for Postwar Unemployment Rate

Symbol	Explanation	MLE	Standard error
c_1	Normal intercept	0.0605	0.0294
c_2	Mild recession intercept	0.307	0.038
c_3	Severe recession intercept	0.715	0.075
ϕ_1	First AR coefficient	0.8584	0.0371
ϕ_2	Second AR coefficient	0.1217	0.0366
σ	Standard deviation/scale	0.134	0.007
ν	Degrees of freedom	5.09	1.01
$[p_{ij}]$	State transition probabilities	$\begin{bmatrix} 0.971 & 0.145 & 0 \\ 0.029 & 0.778 & 0.508 \\ 0 & 0.077 & 0.492 \end{bmatrix}$	$\begin{bmatrix} 0.010 & 0.046 & - \\ 0.010 & 0.065 & 0.193 \\ - & 0.041 & 0.193 \end{bmatrix}$

Friedman (1969, p. 274) and DeLong and Summers (1986), among others, have forcefully advanced the related proposition that asymmetry is the defining characteristic of business cycles. In particular, the asymmetry commented on in Figure 1—that recessions are characterized by an unusually rapid but nonetheless transient increase in the unemployment rate—has been confirmed to be a statistically significant feature of these data in a number of recent quantitative studies, including Montgomery et al. (1998), Rothman (1998), and van Dijk, Franses, and Paap (2002).

Suppose we try in a simple way to represent the asymmetry that the eye perceives in Figure 1. One idea is that the intercept in equation (2) assumes different values in different phases of the business cycle. Consider the following generalization of (2),

$$y_t = c_{s_t} + \phi_1 y_{t-1} + \phi_2 y_{t-2} + u_t, \tag{5}$$

where $s_t = 1$ if the economy is in the normal growth state at date t, $s_t = 2$ if it is in a mild recession, and $s_t = 3$ for a severe recession. Suppose that the transition between these three regimes follows a Markov chain, where $p_{ij} = \Pr(s_t = j \mid s_{t-1} = i)$ is another set of nine parameters to be estimated subject to the constraints $0 \le p_{ij} \le 1$ and $\sum_{j=1}^{3} p_{ij} = 1$. Conditional on the economy being in regime j at date t, the unemployment rate thus has conditional log density given by

$$\log f\left(y_t \middle| s_t = j, y_{t-1}, y_{t-2}, \ldots, y_1\right)$$
$$= k - \left[(\nu+1)/2\right]\log\left[1 + \frac{u_{jt}^2}{\nu\sigma^2}\right]$$
$$u_{jt} = y_t - c_j - \phi_1 y_{t-1} - \phi_2 y_{t-2}$$

for k as in (3).

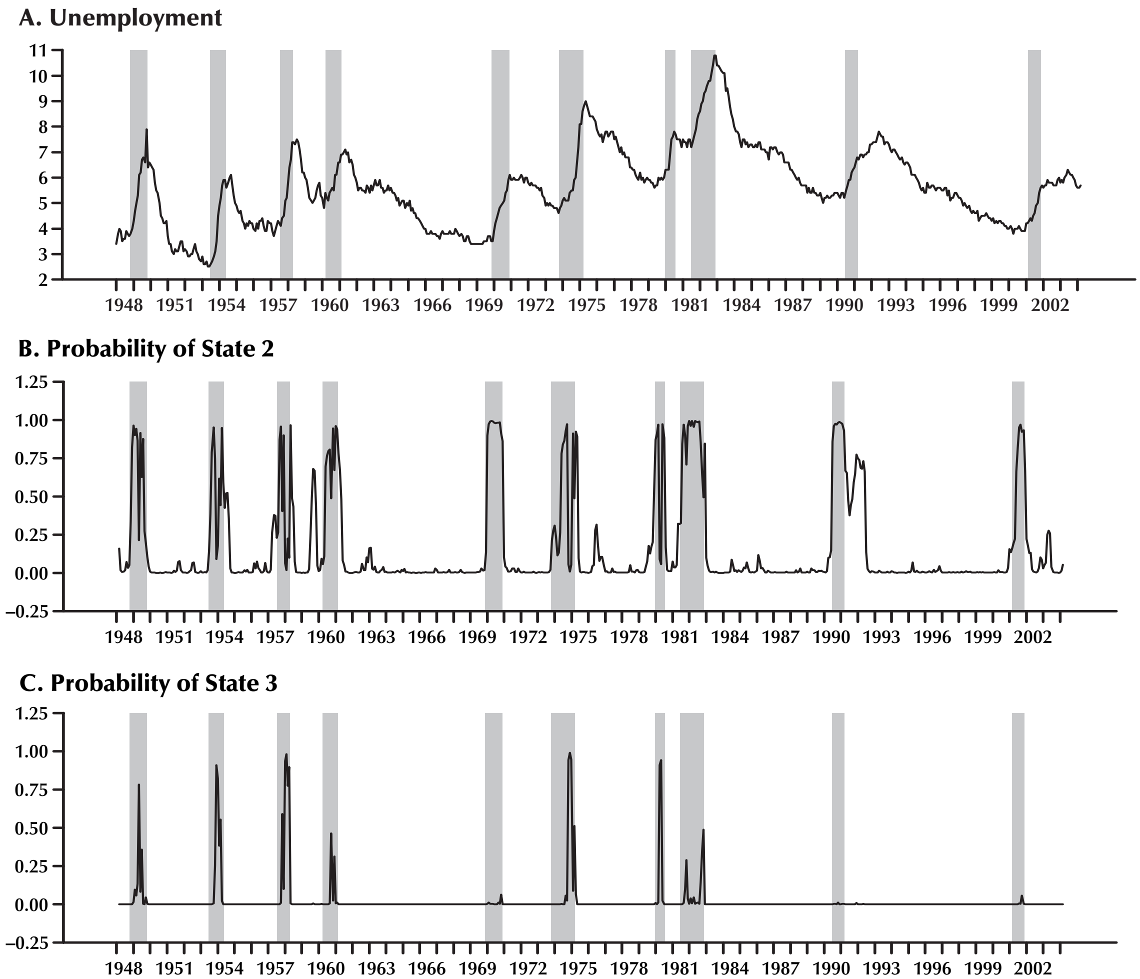
The log likelihood for the observed data,

$$\mathfrak{L}(\theta) = \sum_{t=3}^{T} \log f\left(y_t \middle| y_{t-1}, y_{t-2}, \ldots, y_1; \theta\right),$$

can then be calculated as in Hamilton (1994, equation [22.4.7]) and maximized numerically with respect to the population parameters $\theta = (c_1, c_2, c_3, \phi_1, \phi_2, \sigma, \nu, p_{11}, p_{12}, p_{13}, p_{21}, p_{22}, p_{23}, p_{31}, p_{32}, p_{33})'$. The MLEs turn out to be at the boundaries such that $p_{13} = 0$ and $p_{31} = 0$; that is, states 1 and 3 never follow each other. These MLEs are reported in Table 2, where in order to calculate standard errors from the second derivative matrix of the log likelihood, we went ahead and imposed the constraint $p_{13} = p_{31} = 0$ so that there are just four free transition probabilities (p_{11}, p_{21}, p_{22}, and p_{32}). The smoothed probabilities implied by the MLEs, $\Pr(s_t = j \mid y_1, y_2, \ldots, y_T; \hat{\theta})$, are plotted in Figure 4.

Figure 4

Smoothed Probabilities, Unemployment



NOTE: Panel A reproduces the unemployment data from Figure 1. Panel B is for the moderate recession regime ($\Pr(s_t = 2|y_1,y_2,\ldots,y_T,\hat{\theta})$) and Panel C is for the severe recession regime ($\Pr(s_t = 3|y_1,y_2,\ldots,y_T,\hat{\theta})$) for $\hat{\theta}$ the maximum likelihood estimates reported in Table 2.

Note that allowing this nonlinearity produces a further huge improvement in the log likelihood and would easily be selected on the basis of the Schwarz criterion (again see Table 1). Formal tests of the null hypothesis of no Markov switching have been proposed by Hansen (1992), Hamilton (1996), Garcia (1998), and Carrasco, Hu, and Ploberger (2004). The latter have found the optimal test against the alternative of Markov switching, which we implement here for the special case where the alternative to (4) is a model with Markov switching in the intercept only as in (5); for details of the Carrasco, Hu, and Ploberger test see the appendix. The test statistic on the U.S. employment data turns out to be 26.02, whereas the 5 percent critical value is only 4.01, providing overwhelming evidence for representing the data with the model described in Table 2 rather than (4).

Consider some of the properties of the estimated Markov-switching model. The process is stationary, allowing us to take the unconditional expectation of equation (5),

$$E(y_t)=\sum_{j=1}^{3}c_j \Pr(s_t=j)+\phi_1 E(y_t)+\phi_2 E(y_t)$$

implying

$$E(y_t)=(1-\phi_1-\phi_2)^{-1}\sum_{j=1}^{3}c_j \Pr(s_t=j). \quad (6)$$

The ergodic probabilities can be calculated as in Hamilton (1994, equation [22.2.26]):

$$\begin{bmatrix}\Pr(s_t=1)\\ \Pr(s_t=2)\\ \Pr(s_t=3)\end{bmatrix}=\begin{bmatrix}0.810\\ 0.165\\ 0.025\end{bmatrix}.$$

In other words, the economy spends about 80 percent of the time in the normal phase of the business cycle. Substituting these unconditional probabilities into (6), the model implies an expected unemployment rate of 5.9 percent, close to the postwar average.[4] If the economy is in state i at date t, on average it will stay there for $\sum_{k=1}^{\infty}(1-p_{ii})kp_{ii}^{k-1}=(1-p_{ii})^{-1}$ months. From the transition probabilities in Table 2, the expected duration of each regime is

$$\begin{bmatrix}(1-p_{11})^{-1}\\ (1-p_{22})^{-1}\\ (1-p_{33})^{-1}\end{bmatrix}=\begin{bmatrix}34.0\\ 4.5\\ 2.0\end{bmatrix}.$$

A typical expansion thus might last about three years. A mild recession often lasts less than half a year, at which point it might enter the more severe phase for a few months and then spend another half year again at moderately high unemployment rates before coming back down.

Figure 5 displays the way in which this representation captures the asymmetry of the business cycle discussed here previously. Suppose the economy is currently experiencing an unemployment rate of 6 percent but is in the recovery phase of the cycle ($s_t = 1$). In the absence of any new shocks ($u_{t+j} = 0$ for all j), the dynamic behavior of the unemployment rate over time would be given by

$$y_{t+j}=0.0605+0.8584y_{t+j-1}+0.1217y_{t+j-2}$$

for $j = 1,2,\ldots$ starting from $y_t = y_{t-1} = 6.0$. The top panel of Figure 5 plots this path for y_{t+j} as a function of j. The unemployment rate falls quite gradually in this phase. If the phase persisted indefinitely, the unemployment rate would eventually stabilize at a value of

$$0.0605/(1-0.8584-0.1217) = 3.04,$$

though it would take 30 years of expansion to get there; in the first year of the expansion, the unemployment rate would drop only 0.6 percent.

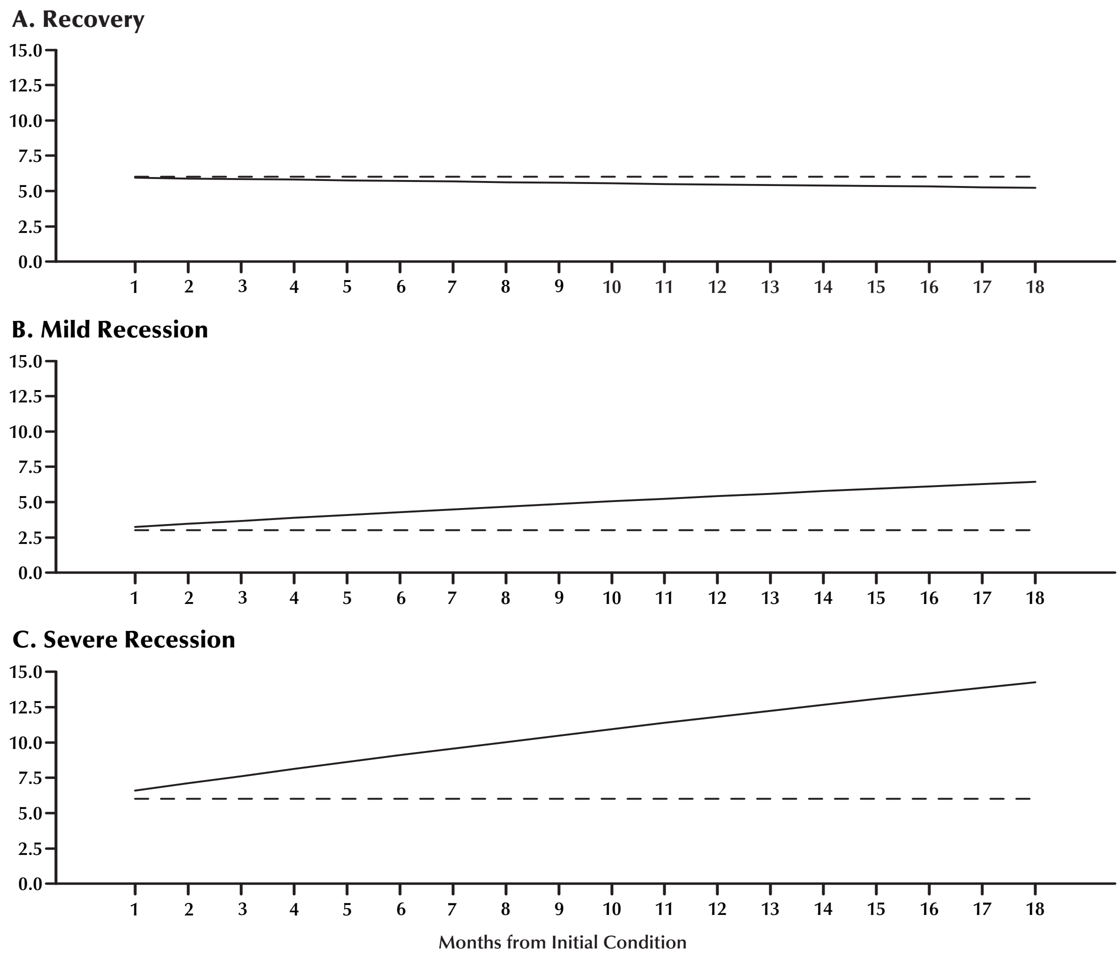
By contrast, if the economy started with $y_t = 3.0$ and entered the mild recession phase, the unemployment rate would rise much more quickly than it fell, reaching 5.4 percent within a year (see Panel B of Figure 5). And if the economy starts with unemployment at 6 percent and enters the severe recession phase (Panel C), the rate would reach nearly 12 percent after a year.

This asymmetry in unemployment dynamics is subtle in the statistical sense that the baseline model (4) adequately captures most of the gross features of the data. But it seems extremely important from an economic perspective. For once we agree that up and down movements in the unemployment rate occur at different speeds, the notion that they may be caused by different economic forces becomes much more appealing. Finding this kind of nonlinearity using modern statistical methods could thereby be interpreted as exonerating to some degree the assumptions of early students of the business cycle. Specifically, it suggests that there is a real event associated with the label of an "economic recession," which is measurable by objective statistical methods as a change in the dynamic behavior of the unemployment rate. It further suggests that the key task for business cycle theorists should not be to look for a unification between the explanation of short-run dynamics and long-run growth, but rather to identify the factors that can result in a temporary failure of the economy to utilize efficiently the available labor, capital, and technology.

[4] Unlike ordinary least squares, the maximum likelihood estimates in this case do not imply an estimate of the population mean that is equal by construction to the sample mean.

Figure 5

Simulated Time Path for Unemployment Under Three Scenarios



NOTE: In Panel A, unemployment is initially at 6 percent and the economy remains in regime 1 with no further shocks; in Panel B, unemployment is initially at 3 percent and the economy remains in regime 2 with no further shocks; in Panel C, unemployment is initially at 6 percent and the economy remains in regime 3 with no further shocks.

THE BEHAVIOR OF INTEREST RATES

I argued above that the business cycle is not an artificial categorization existing only in the imagination of economists but rather is an integral part of the real data-generating process. In addition to believing that such patterns exist, early students of the business cycle also differed from their modern counterparts in the degree to which financial factors were viewed as a key part of the business cycle. For example, Burton (1902) described the typical business downturn in the following terms:

> The usual signal for the beginning of a crisis is a conspicuous banking or mercantile failure, or the exposure of some fraudulent enterprise which attracts wide-spread attention. Money is hoarded. Credit is refused. (Burton, 1902)

As a first look at some of the evidence that may have inspired such an opinion, consider Figure 6,

Figure 6

Monthly Short-Term Commercial Paper Rates in New York City, 1857:01–1937:01

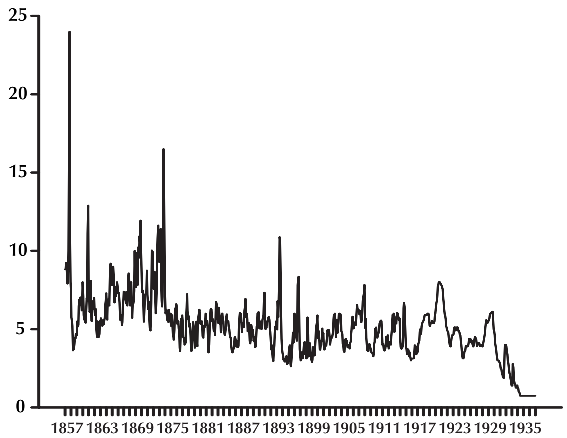
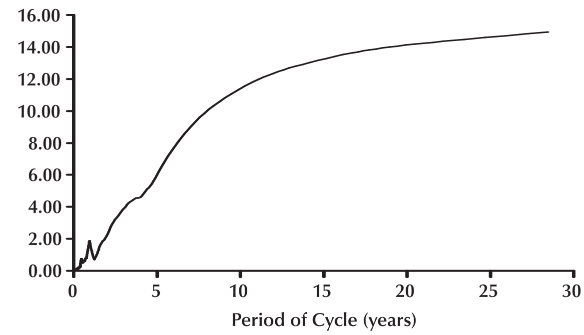


Figure 7

Estimated Spectrum of U.S. Monthly Short-Term Commercial Paper Rates, 1857:01–1913:12



NOTE: Plotted as a function of the period of the cycle in years.

which plots the series on commercial paper rates in New York city compiled by Macaulay (1938).[5] These data behave very differently before and after the founding of the U.S. Federal Reserve in 1913, a point to which we will want to return shortly. But first consider the behavior of this series prior to 1913. There is at least one asymmetry that Figure 1 and Figure 6 have in common, induced per force by the fact that neither the unemployment rate nor the nominal interest rate can ever go negative, but both can go and occasionally have gone quite far above their usual levels. Notwithstanding, interest rates do not display the same obvious pattern of ups and downs as the unemployment data in Figure 1. Indeed, apart from seasonal fluctuations stemming from the predominance of agriculture in the economy over this period, such predictable swings might be difficult to reconcile with the view that asset markets function efficiently given investors' ability to arbitrage between securities of different maturities. Nevertheless, there are a number of interest-rate spikes prior to 1913 that are quite dramatic visually and that have been delegated particular names as important points of reference for economic historians, names such as the Crisis of 1857, Crisis of 1873, or the Panic of 1893.

As in the case of postwar unemployment rates, there is no obvious periodicity to the occurrence of these interest rate spikes. Figure 7 plots the spectrum of interest rates over the 1857:01 to 1913:12 period, which again is dominated by the lowest frequencies. There is a noticeable local peak associated with cycles of a 12-month period, reflecting the tendency of interest rates to be highest in the winter months when agricultural borrowing needs were greatest. But again there is no mass noticeable for what are often described as "business cycle frequencies," namely, cycles with periods of 3 to 5 years.

For a baseline model, we estimated a slight generalization of the model in equations (1) through (3) to allow for a possible seasonal pattern in interest rates, replacing equation (2) with

$$u_t = z_t - \phi_1 z_{t-1} - \phi_2 z_{t-2}$$
$$z_t = y_t - c - \gamma \cos(\pi t / 6) \tag{7}$$

for y_t now the interest rate and $t = 1$ corresponding to the January 1857 observation. A positive value of γ would reflect the tendency for interest rates to peak around December. The likelihood was then

[5] Taken from Macaulay (1938, Column 3, Table 10, pp. A142-61). Macaulay describes these as "choice 60-90 day two name paper" up to 1923 and "four to six month prime double and single name paper" from 1924 on.

Table 3

Maximum Likelihood Estimates of the Baseline Model for Nineteenth-Century Interest Rates

Symbol	Explanation	MLE	Standard error
c	Typical interest rate	5.11	0.18
ϕ_1	First AR coefficient	1.104	0.031
ϕ_2	Second AR coefficient	–0.212	0.026
γ	Effect of seasonal	0.444	0.070
σ	Scale factor	0.407	0.021
ν	Degrees of freedom	2.28	0.24

Table 4

Comparison of Selected Models of Nineteenth-Century Interest Rates

Model	No. of parameters	Log likelihood	Schwarz criterion
Gaussian AR(2)	5	–932.67	–948.99
Student *t* AR(2)	6	–677.05	–696.63
Student *t* AR(2) with MS variance	9	–633.58	–662.95

NOTE: All models also include cosine seasonal. Schwarz criterion calculated as $\mathfrak{L} - (k/2)\log(T)$ for $\mathfrak{L}$ the log likelihood, k the number of parameters, and $T = 682$ the sample size.

maximized numerically over 1857:03 to 1913:12, resulting in the estimates reported in Table 3. Once again, allowing Student *t* innovations implies a huge improvement in the log likelihood over a Gaussian specification (see Table 4).

Apart from the seasonal cycle in agriculture, economic theory would lead us to expect that changes in interest rates are much harder to forecast than the unemployment rate. A Markov-switching specification related to that in (5) is to allow the variance rather than the intercept to change with the regime, so that the conditional log density of the *t*th observation is

$$\log f\left(y_t \middle| s_t = j, y_{t-1}, y_{t-2}, \ldots, y_1\right)$$
$$= k_j - \left[(\nu+1)/2\right]\log\left[1 + \frac{u_t^2}{\nu\sigma_j^2}\right]$$

$$k_j = \log\left\{\Gamma\left[(\nu+1)/2\right]\right\} - \log\left\{\Gamma\left[\nu/2\right]\right\}$$
$$-(1/2)\log\left(\sigma_j^2\nu\pi\right),$$

where u_t is again given by (7). Here we allow only two regimes, with the scale parameter taking the value σ_1 in the low-variance regime and σ_2 in the high-variance regime.

As seen in Table 4, allowing the possibility of changes in the variance parameter over time again leads to a very large improvement in the fit to the data. We can also again use the Carasco, Hu, and Ploberger test, this time constructed for the alternative of a Markov-switching variance, as described in the appendix. The test statistic is 63.34, which vastly exceeds the 5 percent critical value of 3.68, and indeed exceeds the largest value (19.28) among any of our 1,000 Monte Carlo simulations.

The MLEs are reported in Table 5. Regime 2 is characterized by a variance of interest rate innovations that is thrice that for regime 1, with the ergodic probabilities implying that the economy would spend about 80 percent of the time in the low-variance regime. Smoothed probabilistic inferences about when U.S. interest rates were

Table 5

Maximum Likelihood Estimates of Markov-Switching Model for Nineteenth-Century Interest Rates

Symbol	Explanation	MLE	Standard error
c	Typical interest rate	5.11	0.19
ϕ_1	First AR coefficient	1.079	0.040
ϕ_2	Second AR coefficient	–0.178	0.039
γ	Effect of seasonal	0.441	0.064
σ_1	Scale factor in regime 1	0.612	0.019
σ_2	Scale factor in regime 2	1.093	0.092
p_{11}	Probability regime 1 follows itself	0.9867	0.0068
p_{22}	Probability regime 2 follows itself	0.941	0.033
ν	Degrees of freedom	4.40	0.90

unusually volatile are graphed in the middle panel of Figure 8 and repeated in the bottom graph for comparison with NBER-dated recessions (shown as shaded regions) over this period. Those dates for which the smoothed probability of being in the high-variance regime, $\Pr(s_t = 2 \mid y_1, y_2, \ldots, y_T; \hat{\theta})$, exceeds one-half are categorized in Table 6 as episodes of unusually volatile interest rates for further comparison with the dates NBER has ascribed to economic recessions.[6] The correspondence between these two sets of dates is strong, but not perfect. There are a half-dozen recessions in the 1880s and early 1900s, when nothing much was happening to interest rates, and one interest rate spike (in 1898) that comes in the middle of an economic expansion. A long episode of interest rate volatility in the late 1860s and early 1870s also correlates rather weakly with the recessions at those times. On the other hand, there are a number of other recessions, notably those of 1857-58, 1860-61, 1893, and 1896, that match up perfectly with very dramatic shifts in the interest rate regime. The suggestion thus seems strong that an important shift in the interest rate process is indeed related to some but by no means all of the nineteenth-century economic downturns.

Table 6

Dates of Volatile Interest Rates and U.S. Recessions, 1857-1913

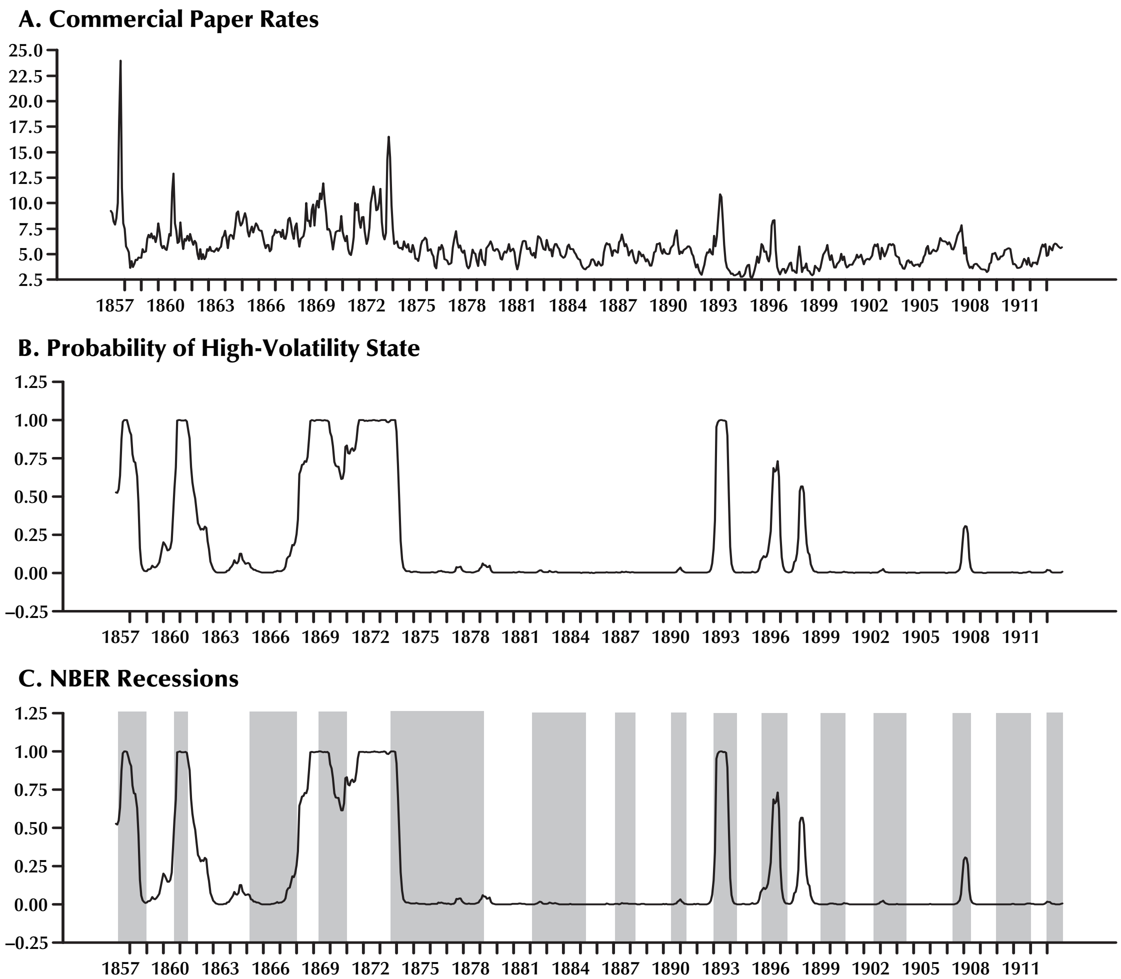
Volatile interest rates	Economic recessions
1857:03–1858:06	1857:06–1858:12
1860:09–1861:11	1860:10–1861:06
—	1865:04–1867:12
1868:03–1874:02	1869:06–1870:02
	1873:10–1879:03
—	1882:03–1885:05
—	1887:03–1888:04
—	1890:07–1891:05
1893:03–1893:12	1893:01–1894:06
1896:07–1896:12	1895:12–1897:06
1898:03–1898:06	—
—	1899:06–1900:12
—	1902:09–1904:08
—	1907:05–1908:06
—	1910:01–1912:01
—	1913:01–1914:12

The asymmetry found in unemployment dynamics—that unemployment rises above its average rate more quickly than it falls below its average rate—seemed to suggest rather directly that macroeconomic theorists should be looking for different explanations as to why employment rises and falls. By contrast, the deficiency of a homoskedastic linear model as a description of nineteenth-century interest rate dynamics—that interest rates are much more volatile at some times

6 These can be found at www.nber.org/cycles/cyclesmain.html.

Figure 8

Smoothed Probabilities, Commerical Paper Rates



NOTE: Panel A shows interest rate data from Figure 6 for 1857:01–1913:12. Panel B shows the smoothed probability for the high-volatility interest rate regime ($\Pr(s_t = 2|y_1,y_2,\ldots,y_T,\hat{\theta})$) for $\hat{\theta}$ the maximum likelihood estimates reported in Table 5. Panel C shows the same series as Panel B, with NBER-dated recessions indicated as shaded areas.

than others—would be consistent with any model that implies GARCH effects on interest rates.[7] However, it does seem fair to describe this finding as implying that, at least during some of the nineteenth-century recessions, interest rates were being influenced by some forces that do not operate in usual times, or, perhaps more strongly, that the financial crises emphasized by early students of the business cycle are in some important respects qualitatively different from the factors governing normal interest rate fluctuations.

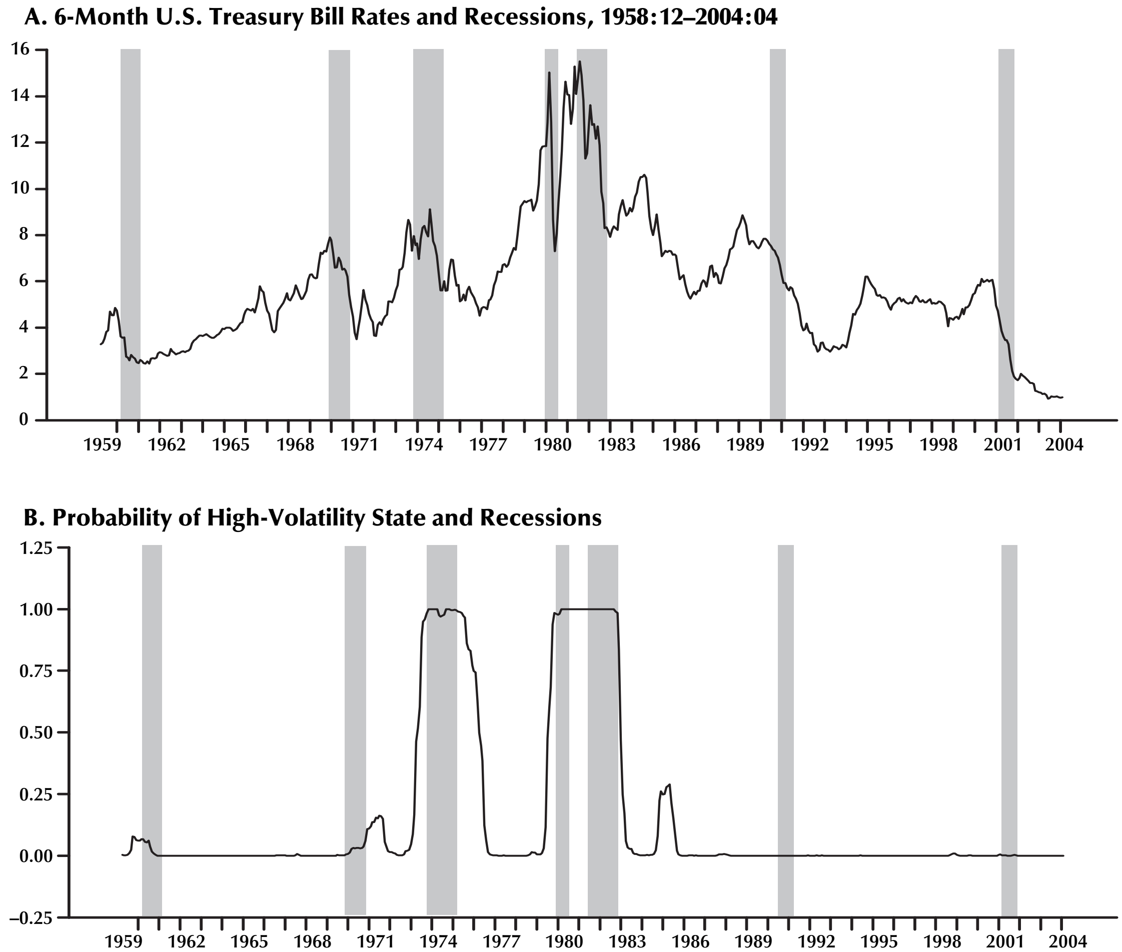
It is quite apparent from Figure 6 that the behavior of short-term interest rates changed dramatically after the Federal Reserve Act in 1913, as both the seasonality and the sharp spikes in interest rates were successfully eliminated.[8] However, a few broad surges in interest rates after the

[7] See, for example, Granger and Machina (2004).

[8] See Miron's (1986) interesting discussion of the possible relation between these two facts.

Figure 9

Smoothed Probabilities, Treasury Rates



NOTE: Panel B shows the smoothed probability for the high-volatility interest rate regime ($\Pr(s_t = 2|y_1, y_2, \ldots, y_T; \hat{\theta})$) for $\hat{\theta}$ the maximum likelihood estimates reported in Table 8. NBER-dated recessions are indicated as shaded areas.

founding of the Fed are certainly identifiable in the post-Fed data. These precede the economic downturn of 1920-21 and the Great Depression beginning in 1929, in addition to a minor spike coinciding with the worsening economic situation in 1931.[9] This post-Fed correlation led many economists to conclude that the contribution of liquidity crunches to economic downturns did not end with the founding of the Federal Reserve.

Indeed, there is some indication that this pattern is a characteristic of postwar business cycles as well. The top panel of Figure 9 displays U.S. 6-month Treasury bill rates[10] from 1958:12 to 2004:04 along with U.S. recessions as shaded areas. Again one's eye is tempted to see a recurrent tendency for the interest rate to surge upward prior to every postwar recession, though the pat-

[9] Among the many who discuss these events are Friedman and Schwartz (1963) and Hamilton (1987, 1988).

[10] These are averages over business days during the month from the secondary market. Data are from the Federal Reserve Economic Database (FRED®II) of the Federal Reserve Bank of St. Louis: http://research.stlouisfed.org/fred2/.

Table 7

Comparison of Selected Models of Changes in Postwar Interest Rates

Model	No. of parameters	Log likelihood	Schwarz criterion
Gaussian AR(2)	4	–292.30	–304.88
Student *t* AR(2)	5	–150.66	–166.38
Student *t* AR(2) with MS variance	8	–87.96	–113.13

NOTE: Schwarz criterion calculated as $\mathfrak{L} - (k/2)\log(T)$ for $\mathfrak{L}$ the log likelihood, k the number of parameters, and $T = 540$ the sample size.

Table 8

Maximum Likelihood Estimates of the Markov-Switching Model for Changes in Postwar Interest Rates

Symbol	Explanation	MLE	Standard error
c	Intercept	0.0016	0.0096
ϕ_1	First AR coefficient	0.409	0.049
ϕ_2	Second AR coefficient	–0.046	0.042
σ_1	Scale factor in regime 1	0.174	0.013
σ_2	Scale factor in regime 2	0.740	0.107
p_{11}	Probability regime 1 follows itself	0.9941	0.0049
p_{22}	Probability regime 2 follows itself	0.9635	0.0247
ν	Degrees of freedom	4.27	0.94

tern is more convincing for some recessions than for others.

Bansal and Zhou (2002), Ang and Bekaert (2002), and Dai, Singleton, and Yang (2003) have used a Markov-switching framework to suggest that a connection between interest rate volatility and economic recessions also characterizes postwar data. We applied the method used to analyze nineteenth-century interest rates to postwar data, with two changes. First, the term representing the agricultural seasonal was dropped. Second, we estimated the model in differences rather than levels. The reason is that levels estimation results in one root that is virtually unity, which rarely introduces numerical problems when maximizing the log likelihood for simulated samples in the Monte Carlo calculation of the Carasco, Hu, and Ploberger statistic (though results to be reported here are otherwise identical for estimation in either levels or differences).

Table 7 replicates the findings of our other two data sets: The data are far better described with Student *t* as opposed to Normal innovations, and further allowing for Markov-switching in the variance yields a further huge improvement in fit. Again the Carrasco, Hu, and Ploberger statistic of 137.38 far exceeds the 95 percent critical value of 4.03 and indeed any value in our 1,000 Monte Carlo simulated samples.

MLEs are reported in Table 8. The standard deviation of the innovation is four times as large in regime 2 compared with regime 1. The bottom panel of Figure 9 plots the smoothed probability that the economy was in the high-volatility regime for each month. As in the nineteenth-century data, there appears to be a clear connection between interest rate volatility and economic recessions, at least for the recessions of 1973-75, 1980, and 1981-82, confirming Bansal and Zhou's (2002), Ang and Bekaert's (2002), and Dai, Singleton, and

Yang's (2003) conclusions from the volatility of the term structure.

Nor is this pattern unique to the United States. For example, Neumeyer and Perri (2004) noted that domestic interest rate spikes preceded the strong economic downturns in Mexico in 1994, Brazil and Argentina in 1995, and Korea and the Philippines in 1997, each of which had important similarities with the financial crises seen repeatedly in the United States in the nineteenth century.

CONCLUSION

We now return to the question posed in the title of this paper, "What's real about the business cycle?" If an accurate statistical description of the dynamic behavior of key economic magnitudes is a time-invariant linear model such as equation (4), then the correct answer is, there is no such thing as a business cycle. In this case, we might as well stop using the phrase "the business cycle," inherited from a less-informed century, and continue with our efforts at trying to describe short-run economic fluctuations as being governed by the same factors that determine long-run economic growth.

However, we've encountered substantial statistical evidence that there are some features of both unemployment and interest rate dynamics that are inconsistent with a time-invariant linear specification. Each of these series seems to exhibit different dynamic behavior in recessions and expansions. The unemployment rate rises more quickly than it falls, and this often occurs at the same time as rapid spikes up and then back down in short-term interest rates. If moving between such episodes is indeed what we mean by "the business cycle," then the correct answer to our question is, the business cycle is very much a real, objectively measurable phenomenon, in which case keeping both the phrase as well as the research program that the expression invites seems well worthwhile.

The paper's title is of course also a play on words, insofar as "real business cycle theory" is derived from a class of models in which short-run economic fluctuations are driven by the same real (i.e., nonmonetary) factors that are responsible for long-run growth. Our findings cast some doubt on the claim that the business cycle is "real" in this sense. The observation that the dynamics change over the course of the business cycle suggests instead that the forces that cause employment to rise may be quite different from those that cause it to fall. Furthermore, the coincidence of this phenomena with large rapid moves in interest rates suggests that something is disrupting the normal functioning of both labor and financial markets. It is difficult to reconcile these facts with a world view in which financial markets play no role in economic fluctuations.

REFERENCES

Ang, Andrew and Bekaert, Geert. "Regime Switches in Interest Rates." *Journal of Business and Economic Statistics*, 2002, *20*(2), pp. 163-82.

Bansal, Ravi and Zhou, Hao. "Term Structure of Interest Rates with Regime Shifts." *Journal of Finance*, 2002, *57*, pp. 1997-2043.

Brock, William A. and Mirman, Leonard J. "Optimal Economic Growth and Uncertainty: The Discounted Case." *Journal of Economic Theory*, 1972, *4*, pp. 479-513.

Burns, Arthur F. and Mitchell, Wesley C. *Measuring Business Cycles*. New York: National Bureau of Economic Research, 1946.

Burton, Theodore E. *Financial Crises and Periods of Industrial and Commercial Depression*. New York: D. Appleton & Co., 1902.

Carrasco, Marine; Liang, Hu and Werner, Ploberger. "Optimal Test for Markov Switching." Working paper, University of Rochester, 2004.

Dai, Qiang; Singleton, Kenneth J. and Yang, Wei. "Regime Shifts in a Dynamic Term Structure Model of U.S. Treasury Bond Yields." Working paper, Stern School of Business, New York University, 2003.

DeLong, J. Bradford and Summers, Lawrence. "Are

Business Cycles Symmetrical?" in Robert J. Gordon, ed., *The American Business Cycle*. Volume 25. Chicago: National Bureau of Economic Research and University of Chicago Press, 1986.

Fisher, Irving. *Booms and Depressions: Some First Principles*. New York: Adelphi, 1932.

Friedman, Milton. *The Optimum Quantity of Money and Other Essays*. Chicago: Aldine, 1969.

Friedman, Milton and Schwartz, Anna J. *A Monetary History of the United States, 1867-1960*. Princeton, NJ: Princeton University Press, 1963.

Garcia, Rene. "Asymptotic Null Distribution of the Likelihood Ratio Test in Markov Switching Models." *International Economic Review*, 1998, *39*, pp. 763-88.

Granger, Clive W.J. and Machina, Mark J. "Structural Attribution of Observed Volatility Clustering." Working paper, University of California, San Diego, 2004.

Hamilton, James D. "Monetary Factors in the Great Depression." *Journal of Monetary Economics*, 1987, *19*, pp. 145-69.

Hamilton, James D. "The Role of the International Gold Standard in Propagating the Great Depression." *Contemporary Policy Issues*, April 1988, *6*(2), pp. 67-89.

Hamilton, James D. *Time Series Analysis*. Princeton, NJ: Princeton University Press. 1994.

Hamilton, James D. "Specification Testing in Markov-Switching Time-Series Models." *Journal of Econometrics*, 1996, *70*, pp. 127-57.

Hansen, Bruce E. "The Likelihood Ratio Test under Non-Standard Conditions." *Journal of Applied Econometrics*, December 1992, *7*(Supplement), pp. S61-82; "Erratum," March/April 1996, *11*(2), pp. 195-98.

Kydland, Finn E. and Prescott, Edward C. "Time to Build and Aggregate Fluctuations." *Econometrica*, 1982, *50*, pp. 1345-70.

Macaulay, Frederick R. *The Movements of Interest Rates, Bond Yields and Stock Prices in the United States since 1856*. New York: National Bureau of Economic Research, 1938.

Miron, Jeffrey A. "Financial Panics, the Seasonality of the Nominal Interest Rate, and the Founding of the Fed." *American Economic Review*, 1986, *76*, pp. 125-40.

Mitchell, Wesley C. *Business Cycles: The Problem and Its Setting*. New York: National Bureau of Economic Research, 1927.

Mitchell, Wesley C. *What Happens During Business Cycles*. New York: National Bureau of Economic Research, 1951.

Montgomery, Alan L.; Zarnowitz, Victor; Tsay, Ruey S. and Tiao, George C. "Forecasting the U.S. Unemployment Rate." *Journal of the American Statistical Association*, 1998, *93*, pp. 478-93.

Neumeyer, Pablo A. and Perri, Fabrizio. "Business Cycles in Emerging Economies: The Role of Interest Rates." NBER Working Paper No. 10387, National Bureau of Economic Research, 2004.

Rothman, Philip. "Forecasting Asymmetric Unemployment Rates." *Review of Economics and Statistics*, 1998, *80*, pp. 164-68.

Stock, James H. "Measuring Business Cycle Time." *Journal of Political Economy*, 1987, *95*, pp. 1240-61.

Stock, James H. "Estimating Continuous Time Processes Subject to Time Deformation: An Application to Postwar U.S. GNP." *Journal of the American Statistical Association*, 1988, *83*, pp. 77-85.

van Dijk, Dick; Franses, Philip Hans and Paap, Richard. "A Nonlinear Long Memory Model, with an Application to U.S. Unemployment." *Journal of Econometrics*, 2002, *110*, pp. 135-65.

APPENDIX

THE CARRASCO, HU, AND PLOBERGER TEST FOR MARKOV SWITCHING

As in equation (1), let $\ell_t(\theta)$ denote the conditional log likelihood of the tth observation under the null hypothesis of no Markov switching, and let $\ell_t^{(1)}(\theta)$ and $\ell_t^{(2)}(\theta)$ denote its first and second derivatives with respect to the intercept c:

(8) $$\ell_t^{(1)}(\theta) = z_t u_t$$

(9) $$\ell_t^{(2)}(\theta) = -z_t + \frac{2z_t^2 u_t^2}{\nu + 1},$$

where

$$z_t = \frac{\nu + 1}{\nu\sigma^2 + u_t^2}.$$

Define

$$\gamma_t\left(\rho;\hat{\theta}\right) = \ell_t^{(2)}\left(\hat{\theta}\right) + \left[\ell_t^{(1)}\left(\hat{\theta}\right)\right]^2 + 2\sum_{s<t} p^{t-s}\ell_t^{(1)}\left(\hat{\theta}\right)\ell_s^{(1)}\left(\hat{\theta}\right),$$

where ρ is a nuisance parameter characterizing the serial correlation of the Markov chain for s_t under the alternative hypothesis of Markov switching. Finally, let $\mathbf{h}_t(\hat{\theta})$ denote the entire vector of derivatives evaluated at the MLE, of which $\ell_t^{(1)}(\hat{\theta})$ is the first element,

$$\mathbf{h}_t\left(\hat{\theta}\right) = \frac{\partial \ell_t(\theta)}{\partial \theta}\bigg|_{\theta=\hat{\theta}} = \begin{bmatrix} z_t \hat{u}_t \mathbf{x}_t \\ -\dfrac{1}{2\hat{\sigma}^2} + \dfrac{z_t \hat{u}_t^2}{2\hat{\sigma}^2} \end{bmatrix},$$

for $\mathbf{x}_t = (1, y_{t-1}, y_{t-2})'$ and let $\hat{\varepsilon}_t(\rho;\hat{\theta})$ denote the residual from an ordinary least squares regression of $(1/2)\gamma_t(\rho;\hat{\theta})$ on $\mathbf{h}_t(\hat{\theta})$.Carrasco, Hu, and Ploberger proposed calculating

(10) $$C\left(\rho;\hat{\theta}\right) = \left[\max\left\{0, \frac{\sum_{t=1}^T \gamma_t(\rho)}{2\sqrt{\sum_{t=1}^T \left[\hat{\varepsilon}_t\left(\rho;\hat{\theta}\right)\right]^2}}\right\}\right]^2$$

and finding the maximum value of $C(\rho;\hat{\theta})$ over a fixed range of alternatives, say, $\rho \in [0.2, 0.8]$. In my applications this was achieved by calculating the value of

$$v\left(\lambda;\hat{\theta}\right) = \frac{\sum_{t=1}^T \gamma_t\left(\rho(\lambda);\hat{\theta}\right)}{2\sqrt{\sum_{t=1}^T \left[\hat{\varepsilon}_t\left(\rho(\lambda)\right);\hat{\theta}\right]^2}}$$

for

$$\rho(\lambda) = 0.2 + 0.6\frac{|\lambda|}{\sqrt{(1+\lambda^2)}}$$

and maximizing $v(\lambda;\hat{\theta})$ numerically with respect to λ.

Critical values for the test statistic ν were obtained by parametric bootstrap as follows. A total of $M = 1{,}000$ samples each of size $T = 673$ were generated from (4). For the mth artificial sample, the MLE $\hat{\theta}^{(m)}$ was found by numerical search for that sample and then the function $\nu(\lambda;\hat{\theta}^{(m)})$ was maximized numerically with respect to λ. The p-value of the observed statistic for the original sample $\nu(\lambda^*,\hat{\theta})$ was then estimated from the fraction of the M samples for which $\nu(\lambda^{*(m)},\hat{\theta}^{(m)})$ exceeded $\nu(\lambda^*,\hat{\theta})$.

To test the null of constant regimes against the alternative of Markov switching in the variance, we replace (8) and (9) with

$$\ell_t^{(1)}(\theta) = -\frac{1}{2\sigma^2} + \frac{z_t u_t^2}{2\sigma^2}$$

$$\ell_t^{(2)}(\theta) = \frac{1}{2\sigma^4} + \frac{z_t u_t^2}{2\sigma^4} - \frac{z_t^2 u_t^2 \nu}{2\sigma^2(\nu+1)}.$$

Commentary

Mark W. Watson

Hamilton's paper (2005) asks two provocative questions. First, are business cycles "real" in the sense that recession/expansion phases represent fundamental shifts in the dynamic model characterizing the macroeconomy? Second, are business cycles "real" in the sense of being caused by real shocks such as productivity or labor supply? Hamilton's careful empirical analysis of the postwar unemployment rate and of interest rates in nineteenth century and postwar periods leads him to answer yes to the first question; on the basis of this analysis he conjectures that the answer to the second question is no.

My comments will address the first of Hamilton's questions. I first ask whether the nonlinear switching models of the sort estimated by Hamilton for the unemployment rate are necessary to explain business cycles of the sort experienced by the United States in the postwar period. My answer, like the answer given by Slutsky (1937), is no. I next ask whether nonlinear models provide a better fit and produce more-accurate forecasts than linear models for postwar U.S. macroeconomic data. My answer is a cautious maybe.

DO WE NEED "REAL" BUSINESS CYCLES TO EXPLAIN THE CYCLE?

The first panel of Figure 1 plots quarterly values of the logarithm of real gross domestic product (GDP) for the United States relative to its value in 1948. Evident in the figure is sustained growth that is occasionally interrupted by one of the ten recessions in the postwar period. These alternating periods of expansion and recession are the "business cycle," and in his paper, Hamilton uses a Markov-switching model for the unemployment rate to delineate these expansions and recessions. As Hamilton shows, the dynamics of the unemployment rate are quite different in the expansion and recession states. In this sense, the business cycle is real; that is, unemployment dynamics are significantly different during expansions and recessions.

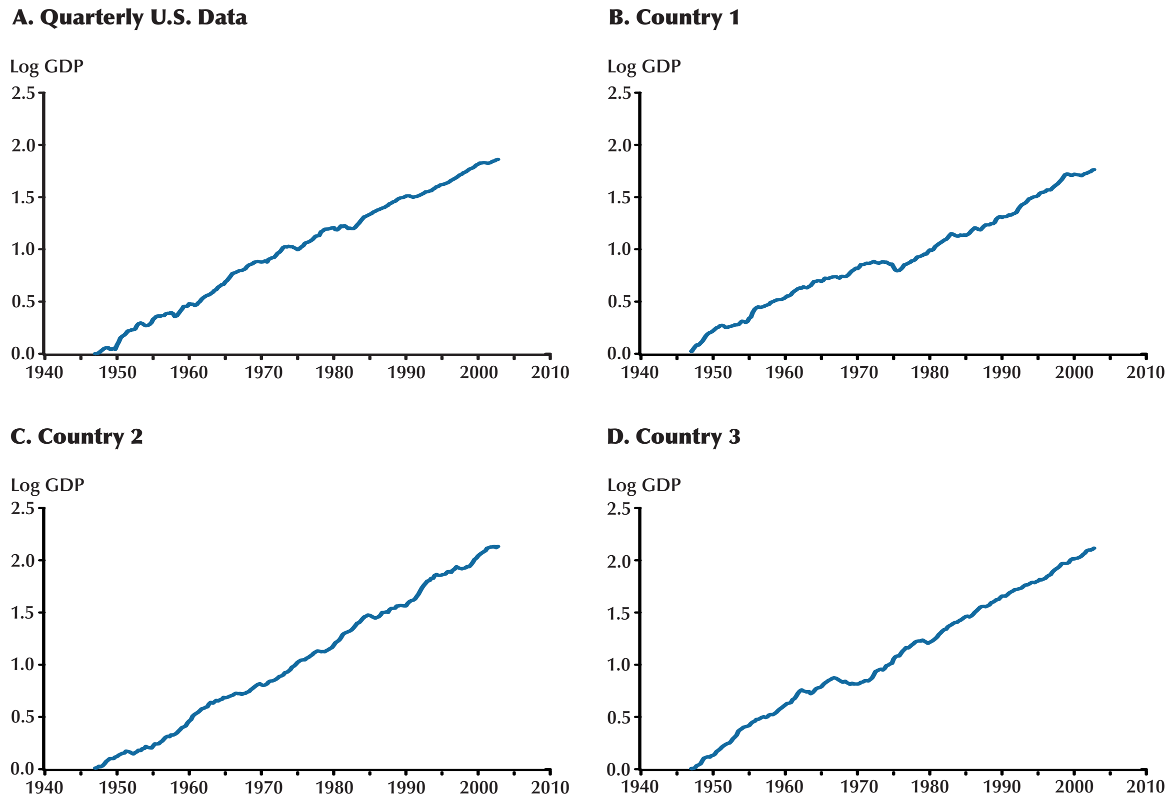
The remaining panels of the figure show results for three other economies, and I often begin my time-series course by asking students to identify the economies that I have plotted. All three countries show periods of expansions and recession like the United States. Country 1 experienced a sharp and severe recession in 1975 along with several other less severe recessions. Country 2 suffered a minor recession in 1953, but then grew more or less steadily until its 1970 recession; it weathered 1975 without a recession, but has suffered four recessions since the late 1970s. Country 3 grew rapidly from 1948 until 1963, when it experienced its first postwar recession, then suffered a prolonged recession from 1967 to 1970, a mild recession in 1980, but has been performing well since then. Foreign students typically have a better idea of international business cycles than domestic students and often recognize these business cycle patterns.

Mark W. Watson is a professor of economics and public affairs at Princeton University and a research associate at the National Bureau of Economic Research. This research was funded in part by the National Science Foundation under grant No. SBR-0214131.

Federal Reserve Bank of St. Louis *Review*, July/August 2005, *87*(4), pp. 453-58.

Figure 1

Logarithm of GDP for the United States and Three Other "Countries"



Most students seem surprised when I announce that I produced the plots for countries 1, 2, and 3 using a random number generator. More precisely, these plots were produced using three realizations from a linear AR(2) model with i.i.d. Gaussian innovations that was calibrated to the U.S. data. Of course, this is just an updated version of the remarkable simulations shown in the classic paper by Slutsky (1937).

Slutsky's simulations have important implications for business cycle analysis. They show that simple time-invariant linear time-series models are capable of generating realizations that have the important cyclical properties that we have come to call the business cycle. Evidently, nonlinear switching models are not required to generate business cycles.

This discussion highlights an important difference in empirical characterizations of the business cycle. One characterization—Hamilton's—is that recessions and expansions represent fundamental shifts in the *stochastic process* characterizing the macroeconomy. Another characterization—Slutsky's—is that recessions and expansions are features of the *realization* of the stochastic process; the process doesn't shift, but sometimes it produces data that decline (recessions) and sometimes it produces data that grow (expansions).

EVIDENCE ON "REAL" BUSINESS CYCLES

While nonlinear models are not required to generate time series with business cycle characteristics, nonlinear models may provide better

Table 1

Performance of Linear and Nonlinear Univariate Forecasting Models for 215 Series

Series category (No. of series)	Percentage best: Linear	Percentage best: Nonlinear
Production (24)	21	79
Employment (29)	21	79
Wages (7)	29	71
Construction (21)	5	95
Trade (10)	50	50
Inventories (10)	30	70
Orders (14)	7	93
Money and credit (21)	57	43
Interest rates (11)	45	55
Producer prices (16)	50	50
Consumer prices (15)	63	37
Consumption (5)	80	20
Other (31)	52	48

NOTE: This table summarizes the forecasting experiment in Stock and Watson (1999) involving linear and nonlinear methods for forecasting 215 series in the 13 categories, shown in the first column. The second and third columns show the percentage of series for which the linear model outperformed the nonlinear model (column 2) or the reverse (column 3).

descriptions of the stochastic processes characterizing typical macroeconomic series than simple linear models. A series of papers building on Hamilton's (1989) original contribution have shown that Markov-switching models provide an improvement on the fit of linear models for several important macroeconomic time series. (Examples include Chauvet, 1998, Diebold and Rudebusch, 1996, and Filardo, 1994.) Stock and Watson (1999) compared the forecast performance of various linear and nonlinear univariate forecasting models for 215 monthly macroeconomic time series using a pseudo out-of-sample forecasting experiment. Table 1 contains a summary of their findings. For several categories of series (production, employment, construction, inventories, orders, interest rates, and wages), the nonlinear models outperformed linear models.

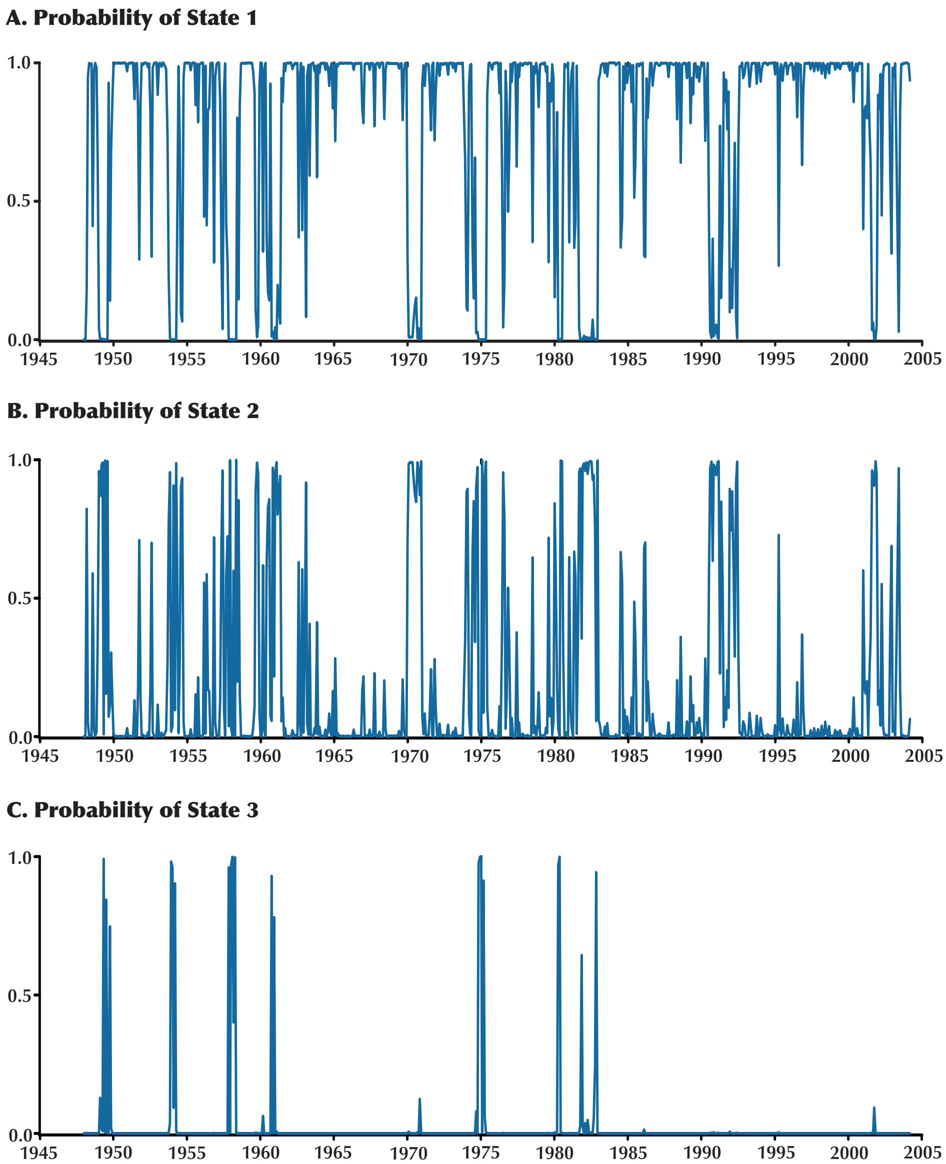
Does the three-state Markov-switching model that Hamilton proposes in this paper produce more precise forecasts of the state of the business cycle than linear models? To investigate this, I computed the one-sided ("filtered") estimates of the state probabilities from Hamilton's model. These filtered probabilities are plotted in Figure 2 and are the one-sided versions of the probabilities plotted in Hamilton's Figure 4. I considered six different monthly series that serve as coincident indicators of the business cycle: the unemployment rate, the index of industrial production (logarithm), real personal income (logarithm), manufacturing and trade sales (logarithm), employment (logarithm), and an index of coincident indicators constructed as a weighted average of the last four series. Using data from 1959-2003, I estimated regression models of the form

(1)
$$y_{t+h} - y_t = \beta_0 + \phi(L)\Delta y_t + \gamma(L)u_t + \beta_2 p_{2t/t} + \beta_3 p_{3t/t} + \varepsilon_{t+h}$$

where y_t denotes the indicator being forecast, u_t denotes the unemployment rate, and $p_{2t/t}$ and $p_{3t/t}$ denote the filtered state probabilities (that is, the nonlinear functions of the unemployment rate plotted in Figure 2). Results for this regression are shown in Table 2 for 1-month-ahead (h = 1)

Figure 2

Filtered Probabilities of Unemployment Rate States



and 3-month-ahead (h =3) forecasts. There does seem to be evidence that $p_{2t/t}$ and $p_{3t/t}$ help predict the indicators, particularly at the 1-month horizon.

An alternative, and arguably more compelling test of the predictive power of $p_{2t/t}$ and $p_{3t/t}$, comes from using recursive estimates of the parameters of (1) to compute pseudo out-of-sample forecasts. Table 3 summarizes results from these calculations over the 1970-2003 out-of-sample forecast period. The results presented in the table are the mean-squared forecast errors for various versions of (1) relative to a simple AR model. The results shown in the column labeled "P" are for the model that includes $p_{2t/t}$ and $p_{3t/t}$ in addition to lags of Δy_t (so that $\gamma(L) = 0$); the results in the column labeled

Table 2

Granger-Casualty Tests for the Model

$y_{t+h} - y_t = \beta_0 + \phi(L)\Delta y_t + \gamma(L)u_t + \beta_2 p_{2t/t} + \beta_3 p_{3t/t} + \varepsilon_{t+h}$

	Forecast horizon $h = 1$			Forecast horizon $h = 3$		
Series forecast	β_2	β_3	F-statistic, p-value	β_2	β_3	F-statistic, p-value
Unemployment rate	0.09 (0.04)	0.34 (0.12)	0.01	0.15 (0.10)	0.35 (0.27)	0.33
Industrial production	–1.08 (2.20)	–11.56 (4.89)	0.05	–1.33 (2.47)	–6.71 (4.00)	0.18
Personal income	–1.39 (1.14)	–2.43 (2.83)	0.47	–1.09 (1.06)	–1.11 (1.85)	0.58
Manufacturing and trade sales	1.15 (2.69)	–11.98 (6.30)	0.05	–1.53 (2.06)	–5.33 (4.60)	0.50
Employment	–0.55 (0.54)	–3.17 (1.38)	0.07	–0.58 (0.56)	–1.47 (1.06)	0.38
Coincident index	–0.94 (1.41)	–6.47 (3.25)	0.13	–1.19 (1.51)	–3.87 (2.74)	0.37

NOTE: The table shows OLS estimates of β_2 and β_3, HAC standard errors, and p-values for the F-test that $\beta_2 = \beta_3 = 0$.

Table 3

Out-of-Sample Mean-Squared Error Relative to Univariate Autoregressive Model

$y_{t+h} - y_t = \beta_0 + \phi(L)\Delta y_t + \gamma(L)u_t + \beta_2 p_{2t/t} + \beta_3 p_{3t/t} + \varepsilon_{t+h}$

	All			Recessions			Expansions		
Series	P	U	U-and-P	P	U	U-and-P	P	U	U-and-P
A. Forecast horizon $h = 1$									
Unemployment	0.98			0.97			0.98		
Industrial production	0.95	0.97	0.97	0.87	0.88	0.88	1.00	1.03	1.02
Personal income	0.95	0.93	0.93	0.92	0.92	0.92	0.95	0.93	0.94
Manufacturing and trade sales	0.99	0.99	1.01	1.01	0.95	1.07	0.99	1.00	0.99
Employment	0.96	0.96	0.97	0.89	0.87	0.86	1.00	1.02	1.02
Coincident index	0.97	0.97	0.97	0.94	0.94	0.95	1.02	1.01	1.02
B. Forecast horizon $h = 3$									
Unemployment	1.01			1.00			1.01		
Industrial production	0.98	0.97	0.98	0.94	0.94	0.95	1.02	1.01	1.02
Personal income	0.92	0.89	0.90	0.90	0.90	0.90	0.93	0.89	0.90
Manufacturing and trade sales	0.99	0.97	0.99	1.00	0.96	1.00	0.99	0.97	0.99
Employment	0.96	0.94	0.96	0.89	0.91	0.90	1.04	0.98	1.02
Coincident index	0.96	0.94	0.95	0.93	0.92	0.93	0.99	0.96	0.98

NOTE: The table shows the mean-squared forecast error for each model relative to that for the univariate AR model. The model labeled "P" imposes the constraint that $\gamma(L) = 0$; the model labeled "U" imposes the constraint that $\beta_2 = \beta_3 = 0$, and the model labeled "U-and-P" imposes no constraints. Results are shown for the full 1970-2003 out-of-sample and for the recession and expansion subsamples.

"U" are for the model that includes lags of u_t and Δy_t (so that $\beta_2 = \beta_3 = 0$), and the results in the column labeled "U and P" include all of the terms in (1). Relative mean-squared errors are shown for the entire out-of-sample period and for recessions and expansions separately.

Most of the table entries are less than 1.0, indicating an improvement on the univariate AR model. However, it is less clear whether the model with nonlinear functions of the unemployment rate ($p_{2t/t}$ and $p_{3t/t}$) outperform the linear model that includes the unemployment rate. There are few entries in which the P or U-and-P models outperform the U model.

My interpretation of the evidence in Tables 2 and 3 is that they provide some additional (albeit weak) evidence supporting the "real" switching model proposed by Hamilton.

REFERENCES

Chauvet, Marcelle. "An Econometric Characterization of Business Cycle Dynamics with Factor Structure and Regime Switches." *International Economic Review*, November 1998, *39*(4), pp. 969-96.

Diebold, Francis X. and Rudebusch, Glenn D. "Measuring Business Cycles: A Modern Perspective." *Review of Economics and Statistics*, February 1996, *78*(1), pp. 67-77.

Filardo, Andrew. "Business-Cycle Phases and Their Transitional Dynamics." *Journal of Business and Economic Statistics*, July 1994, *12*(3), pp. 299-308.

Hamilton, James D. "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle." *Econometrica*, 1989, *57*(2), pp. 357-84.

Hamilton, James D. "What's Real About the Business Cycle?" Federal Reserve Bank of St. Louis *Review*, July/August 2005, *87*(4), pp. 435-52.

Slutsky, Eugen. "The Summation of Random Causes as the Source of Cyclic Processes." *Econometrica*, April 1937, *5*(2), pp. 105-46.

Stock, James H. and Watson, Mark W. "A Comparison of Linear and Nonlinear Univariate Models for Forecasting Macroeconomic Time Series," in Robert F. Engle and Halbert White, eds., *Festschrift in Honor of Clive W.J. Granger*. Oxford: Oxford University Press, 1999, pp. 1-44.

Trends in Hours, Balanced Growth, and the Role of Technology in the Business Cycle

Jordi Galí

This paper revisits a property embedded in most dynamic macroeconomic models: the stationarity of hours worked. First, the author argues that, contrary to what is often believed, there are many reasons why hours could be nonstationary in those models, while preserving the property of balanced growth. Second, the author shows that the postwar evidence for most industrialized economies is clearly at odds with the assumption of stationary hours per capita. Third, he examines the implications of that evidence for the role of technology as a source of economic fluctuations in the G7 countries.

Federal Reserve Bank of St. Louis *Review*, July/August 2005, *87*(4), pp. 459-86.

1 INTRODUCTION

Business cycles have long been associated with highly procyclical fluctuations in labor input measures. In the mind of the common man, the recurrent ups and downs in employment (or unemployment) observed in modern economies are arguably more of a defining feature of the business cycle than the accompanying fluctuations in gross domestic product (GDP).[1] Understanding the factors underlying the joint variation of output and labor input measures remains a key item in macroeconomists' research agenda.

This paper focuses on a dimension of those joint fluctuations that is generally ignored by macroeconomists, in theoretical as well as in empirical analysis: the long-run behavior of hours worked. In particular, the paper revisits a property common to the majority of intertemporal equilibrium models used in macroeconomic applications, namely, that of stationarity of hours worked per capita. First, I argue that, contrary to what is often believed, stationarity of (per capita) hours is not a necessary condition for those models to generate a balanced-growth path. Second, and perhaps most importantly, I show that the evidence for the G7 economies is generally at odds with the key equilibrium relationship that underlies the stationarity of hours in those models. In fact, that evidence suggests that both margins of labor utilization (i.e., hours per worker and the employment rate) display some nonstationarity features in most G7 countries.

The evidence of nonstationarity in hours per worker points to the importance of using an hours-based measure of productivity when estimating the effects of technology shocks under the approach proposed in Galí (1999), in which technology shocks are identified as the only source of nonstationarity in labor productivity. The reason is straightforward: Shocks unrelated to technology that have a permanent effect on hours per worker (but not on output per hour) would be a source of nonstationarity in employment-based measures

1 Notice also that employment is one of the four monthly indicators monitored by the National Bureau of Economic Research's Business Cycle Dating Committee for the dating of recessions.

Jordi Galí is director of and senior researcher at the Centre de Recerca en Economia Internacional (CREi) and a professor of economics at Universitat Pompeu Fabra; he is also a research associate at the National Bureau of Economic Research. The author thanks Chris Sims for useful comments and Anton Nakov for excellent research assistance.

of productivity and would thus be mislabeled as technology shocks. I revisit here the international evidence on the effects of technology shocks using an hours-based measure of productivity and find little evidence in support of a major role of technology as a source of business cycles.

The paper is organized as follows. Section 2 discusses the relationship between the stationarity of hours and the balanced-growth hypothesis. Section 3 provides some evidence on the behavior of the two margins of labor utilization for the G7 countries and discusses the implications of that evidence. Section 4 presents new estimates of the effects and role of technology shocks in the G7 countries.

2 THE STATIONARITY OF HOURS AND THE BALANCED-GROWTH HYPOTHESIS

Since the seminal work by Kydland and Prescott (1982) and Prescott (1986), most business cycle models have adopted a neoclassical growth framework (augmented with a consumption-leisure choice) as a "core structure," on which stochastic disturbances and frictions of different sorts are added. The choice of a specification for preferences and technology in the resulting models has been generally guided by the requirement that the underlying deterministic model is consistent with some "stylized facts" of growth. It is generally argued that imposing such a requirement facilitates calibration of the model on the basis of information unrelated to the business cycle phenomena that the model seeks to explain.

Prominent among those stylized facts is the observation that many key macroeconomic variables such as output, consumption, investment, and the stock of physical capital display a similar average rate of growth over sufficiently long periods of time. That property is referred to as "balanced growth." Another important observation is that hours worked per capita do not display any obvious trend that one could associate with the secular upward trend shown by the real wage. Third, and in contrast to the real wage, the return to capital (as reflected, say, in the real interest rate) does not seem to display any significant long-run trend.

The analysis of King, Plosser, and Rebelo (1988), among others, pointed out that the predictions of the neoclassical framework can be reconciled with the previous facts if (i) technology can be represented by a constant returns production function with labor-augmenting technical progress, (ii) preferences display a constant elasticity of intertemporal substitution, and (iii) the marginal rate of substitution between consumption and hours (or leisure) is homogeneous of degree one in consumption. Specifications satisfying those properties are commonly adopted "as a discipline device" by business cycle theorists, even when their subject of inquiry is viewed as being unrelated to the forces underlying long-term growth.

Unfortunately, as I will argue, the widespread adoption of an unnecessarily strong version of the balanced-growth hypothesis has led to a common misconception—namely, that stationarity of hours worked is an inherent feature of models displaying balanced growth.[2]

2.1 A Benchmark Framework with No Frictions

To illustrate the basic point, let us first assume a constant returns Cobb-Douglas technology, implying the following linear expression for the (log) marginal product of labor[3]:

$$mpn_t = y_t - n_t,$$

where y_t denotes output and n_t is hours worked (or "hours," for short), both normalized by the size of the working-age population and expressed in logs.

Second, and in a way consistent with the requirements derived by King, Plosser, and Rebelo (1988), let us assume that preferences imply the following expression for the (log) marginal rate of substitution:

[2] Cooley and Prescott (1995), among others, provide an explicit account of that strategy: "[W]e are going to restrict our attention to artificial economies that display balanced growth. In balanced growth, consumption, investment, and capital all grow at a constant rate, while hours stay constant."

[3] Constant terms are ignored throughout the paper.

$$mrs_t = c_t + \varphi n_t,$$

where c_t denotes the log of per capita consumption and φ is the reciprocal of the Frisch labor supply elasticity.[4] Notice that normality of both consumption and leisure requires that $\varphi > 0$.

In a benchmark real business cycle (RBC) model with perfect competition in goods and labor markets and no other distortions, the efficiency condition $mrs_t = mpn_t$ holds at all times, implying

$$n_t = -(1 + \varphi)^{-1} s_{c,t}, \tag{1}$$

where $s_{c,t} \equiv c_t - y_t$ is the log of the share of consumption in output (henceforth, the "consumption share"). The intuition behind the negative relationship is simple: Starting from an efficient allocation, both an increase in consumption (given output) or a decline in output (given consumption) make an additional unit of leisure more valuable than the marginal use of time in productive activities, thus calling for a drop in hours to maintain efficiency.

The "strong" version of the balanced-growth property adopted in most macroeconomic applications requires that $s_{c,t}$ fluctuate about a constant mean value s_c. As a result, it must also be the case that hours are stationary and fluctuate around a constant mean $n = -(1 + \varphi)^{-1} s_c$. In the standard RBC model (closed economy, no government), the steady-state consumption share s_c and, hence, steady-state hours, are determined exclusively by preference and technology parameters.

2.2 Hours Worked and the Consumption Share: Evidence for the United States

Next I turn to the data, to provide an assessment of the relationship between hours and the consumption share displayed in (1) and implied by the simple framework noted here previously.

I start by looking at aggregate quarterly U.S. data.[5] I use hours of all persons in the nonfarm business sector (LXNFH), normalized by the size of the population aged 16 or older, as a benchmark measure of hours. As a benchmark series for the consumption share, I use the ratio of personal consumption expenditures in nondurable goods and services (CN+CS) to GDP, both expressed in current prices.[6] Figure 1A displays the empirical counterparts to n_t and $s_{c,t}$. While some negative short-run comovement between the two series is easily discernible, it is clear that the overall picture is dominated by what looks like a common upward trend in the second half of the sample period.

Figure 1B displays the business cycle component of the two series, obtained by applying the band-pass filter developed in Baxter and King (1999) and calibrated to remove fluctuations of periodicity outside an interval between 6 and 32 quarters. Once the low-frequency trends are dispensed with, a strong negative comovement between hours and the consumption share emerges clearly. That relationship is consistent with two stylized facts of business cycles—namely, the procyclicality of hours worked and that consumption is less volatile than output at business cycle frequencies (which makes the consumption share countercyclical).

The previous visual assessment is confirmed by a straightforward statistical analysis. Thus, a simple ordinary least squares (OLS) regression of (log) hours on the (log) consumption share yields the following estimated equation (standard errors in parentheses, constant term ignored):

$$n_t = \underset{(0.05)}{0.16}\, s_{c,t} \qquad R^2 = 0.03,$$

with the simple correlation between the two variables being 0.19. Notice that the fit of the estimated regression equation, as measured by the R^2, is extremely low for a relationship that is supposed to hold exactly. To make things worse, the sign of the estimated coefficient is inconsistent with the theory's prediction.

By way of contrast, an OLS regression using the business cycle component of both series yields the estimated equation

[4] The assumption of a constant elasticity of the marginal rate of substitution with respect to hours can be relaxed, but is adopted here for convenience.

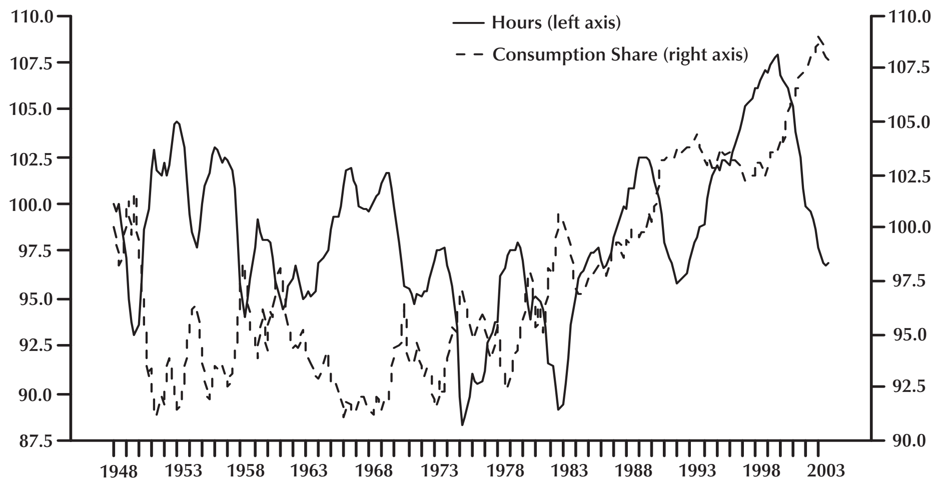
[5] Quarterly U.S. data are drawn from the USECON data set. The sample period is 1948:Q1–2003:Q4. The corresponding mnemonics are shown in parentheses.

[6] In the appendix, I show that this is the right measure if one allows for changes in the relative price of consumption goods.

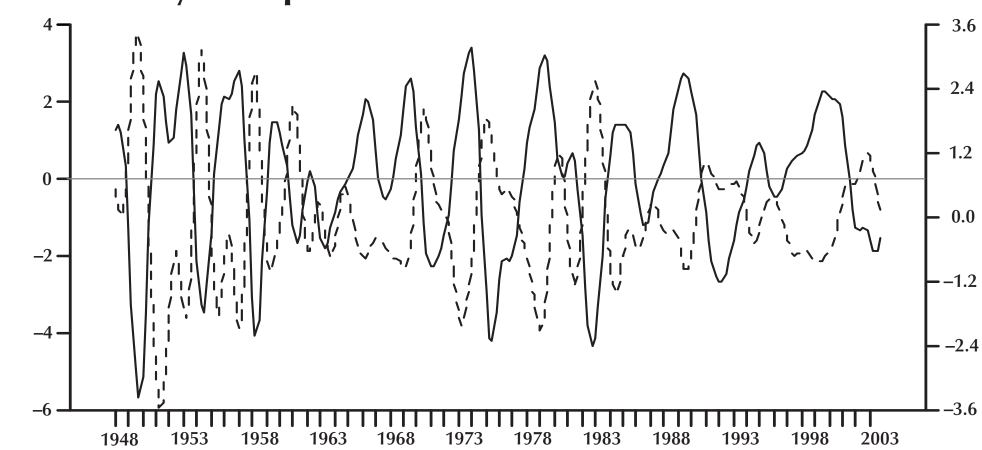
Figure 1

Hours and the Consumption Share: U.S. Evidence
Nondurables and Services

A. Original Data



B. Business Cycle Component



$$n_t = -1.26\, \underset{(0.06)}{s_{c,t}} \qquad R^2 = 0.61,$$

with the corresponding correlation being –0.78. Thus, and while far from displaying the exact relationship implied by the benchmark model, the business cycle component of hours and the consumption share shows a very strong negative comovement. Notice also that, strictly speaking, the estimated coefficient on the consumption share with an absolute value greater than 1 is inconsistent with the theory (since it would imply that leisure is an inferior good). Given the lack of a structural interpretation of the error term, combined with the likely distortions introduced by the detrending filter, I do not want to put too much emphasis on the point estimate.[7] Yet, it is worth noticing that the previous finding seems consistent with the requirement stressed by RBC theo-

[7] A similar finding, albeit in the context of the estimation of a staggered price setting model, is found in Sbordone (2000).

Table 1

Comovement of Hours and the Consumption Share: U.S. Evidence

	Log-levels		Business cycle component	
	s_c	ρ	s_c	ρ
Nonfarm business hours				
Nondurables and services, private	0.16 (0.05)	0.19	–1.26 (0.06)	–0.78
Total private consumption	0.25 (0.06)	0.25	–1.32 (0.08)	–0.71
Total consumption	0.01 (0.04)	0.04	–1.14 (0.05)	–0.76
Total hours				
Nondurables and service, private	0.84 (0.06)	0.66	–1.28 (0.05)	–0.85
Total private consumption	1.00 (0.07)	0.68	–1.17 (0.08)	–0.67
Total consumption	0.69 (0.04)	0.71	–1.13 (0.04)	–0.74

rists of a high intertemporal substitution in labor supply in order to account for the large fluctuations in hours.[8]

2.2.1 Robustness: Alternative Measures. Similar findings emerge when I use an alternative, more comprehensive, measure of *aggregate hours* constructed by multiplying total civilian employment (LE) by average weekly hours in manufacturing (LRMANUA). The same is true for measures of the consumption share constructed using two alternative definitions of consumption, namely, *total private consumption* (C)—which includes durable goods expenditures—and *total consumption* (C+GFNE+GSE)—which includes, in addition, nondefense government consumption.

Table 1 reports the estimates of the coefficient of an OLS regression of (log) hours on the (log) consumption share, using all possible combinations of measures of both variables, and with standard errors reported in parentheses. The second column reports the corresponding correlations. The third and fourth columns show analogous statistics using the business cycle component of each series. Interestingly, when the (unfiltered) data are used, the regression coefficient has the wrong sign for all specifications. When I use data on total hours (see bottom panel), the correlation becomes higher and significant, but always with the wrong sign (positive). The latter just captures the fact that both series display a common upward trend for most of the period.

Again, when we turn to the business cycle component, the results change dramatically: A very strong negative comovement between hours and the consumption share emerges for all the specifications.

2.3 Hours Worked and the Consumption Share: International Evidence

Next I examine the comovements of hours and the consumption share for the G7 countries, in order to assess the extent to which the evidence in the previous subsection is specific to the United States or carries over to other countries. In doing so I use the data set on hours worked (normalized by population aged 14 to 65) constructed by the Organisation for Economic Co-operation and Development (OECD) and part of their Labor Force Statistics data set. The data frequency is annual and the sample period, common across countries, starts in 1970 and ends in 2002. The graphs on the left-hand side of Figure 2 display, for each country, the time series for hours and the share of (total) consumption in GDP. The graphs on the right-hand side show the growth rates of the same variables.[9] That graphical evidence is supplemented with

[8] As exemplified by the indivisible model of Hansen (1985) and Rogerson (1988).

[9] Given the annual frequency of the data, growth rates provide a better representation of short-term fluctuations.

Figure 2

Hours and the Consumption Share: International Evidence

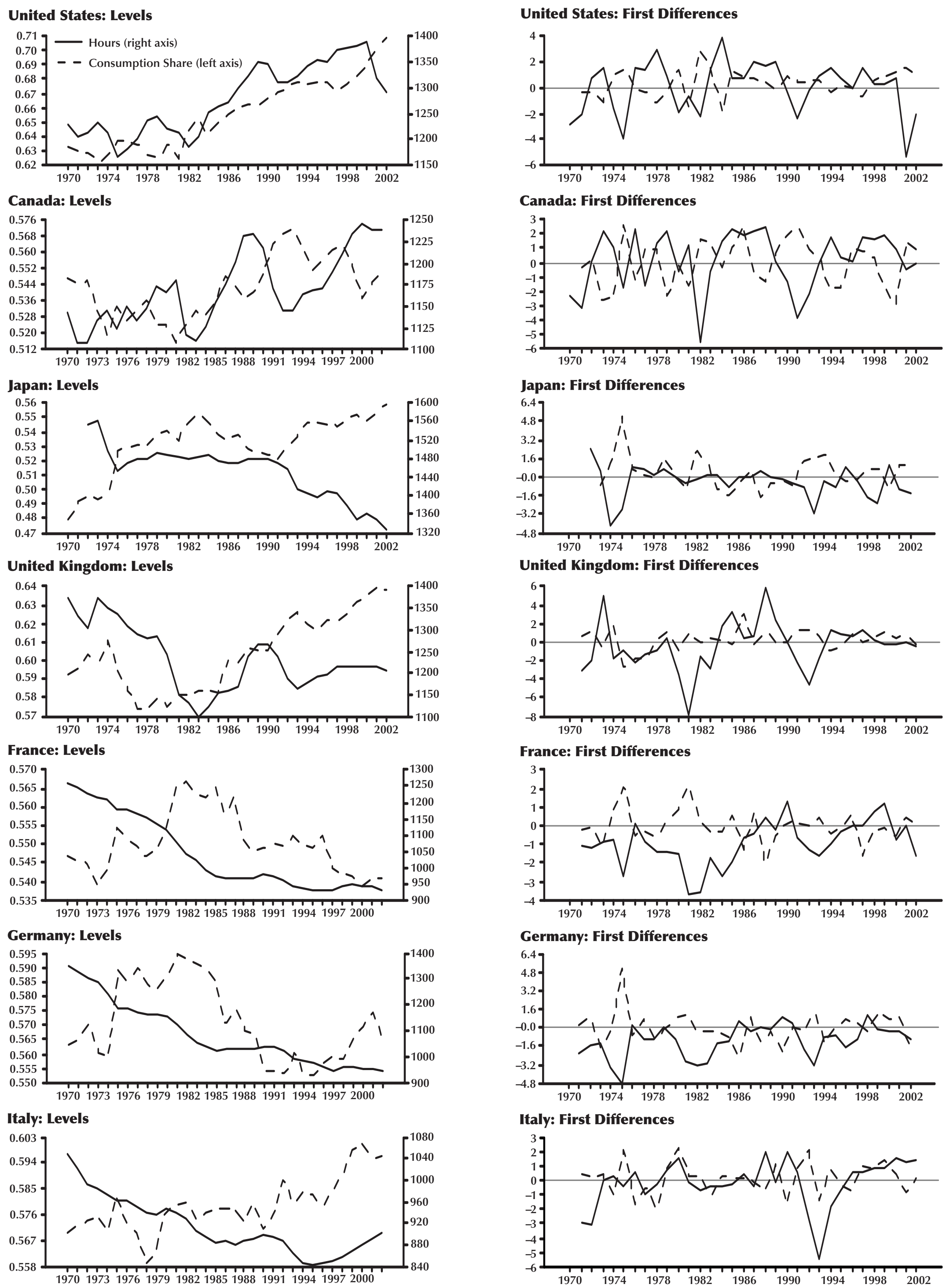


Table 2, which reports, for each country, the correlation between the (log) consumption share and (log) hours, in both levels and first differences.

While the patterns of both variables seen in the figures display substantial differences across countries, two basic common features are apparent. First, neither the consumption share nor hours per capita display any tendency to revert to some constant mean, over the 30-year period considered. In other words, there seems to be prima facie evidence of some sort of nonstationarity in both series. Second, and as confirmed by the correlations in Table 2, there is no evidence of the tight negative relationship between those variables suggested by (1). Japan is the only country for which a strong negative correlation can be found between the (log) levels of the two variables.

As was the case with quarterly U.S. data, when we turn our attention to higher-frequency changes (as represented by first differences here), the correlations become negative (with the exception of Italy), but they are still rather low in absolute value and far from the perfect correlation implied by the benchmark framework above.

2.4 Interpretation

Notice that even under the baseline neoclassical framework described here previously, one can think of plausible reasons that would render the consumption share nonstationary and that could thus provide a potential explanation for the nonstationarity in hours *within the paradigm*. The presence of permanents shifts (or an underlying trend) in the share of government purchases in GDP may be the most obvious one—one that seems to be relevant in the case of the postwar U.S. economy. (From the early 1950s to the late 1990s that share declined by about 5 percentage points.) Yet, according to the benchmark neoclassical framework, the increase in the consumption share should have come hand in hand with a reduction in hours, not an increase like the one observed in the U.S. economy—a pattern displayed also by Canada and Germany—and not with the largely disconnected long-run pattern of hours displayed by other countries (with the possible exception of Japan).

Table 2

Comovement of Hours and the Consumption Share: International Evidence

	Correlation between n and s_c	
	Levels	First differences
United States	0.80	–0.57
Canada	0.23	–0.53
Japan	–0.73	–0.49
United Kingdom	–0.09	–0.10
France	–0.07	–0.47
Germany	0.26	–0.53
Italy	–0.41	0.01

SOURCE: Data are from the OECD Labor Force Statistics Database and Annual National Income Accounts.

More generally, modern economies are subject to a variety of frictions and distortions that can account for permanent shifts—and, thus, unit-root-like behavior—in hours. Furthermore, this can occur without violating the central element of the balanced-growth hypothesis, namely, that over the long run the main components of aggregate demand are expected to grow at the same rate (or, in other words, their growth rates have a common unconditional expectation).

To illustrate this point consider the following extension of the framework above, along the lines of Galí, Gertler, and López-Salido (2003). First, one may want to allow for a possible wedge between the marginal product of labor (which, for simplicity, we keep equating to average labor productivity) and the real wage paid by the firm per hour of work. Letting μ_t^p denote that wedge, we thus have (in logs)

$$w_t - p_t = mpn_t - \mu_t^p = (y_t - n_t) - \mu_t^p. \tag{2}$$

Second, a wedge may exist between the wage and the (average) household's marginal rate of substitution, as a result of labor income taxes, wage setting by unions, efficiency wages, etc.[10] Formally, and letting μ_t^w denote that wedge, we have

[10] This is the wedge emphasized in Mulligan (2002), who provides an analysis of its long-run behavior in the United States.

$$w_t - p_t = mrs_t + \mu_t^w. \tag{3}$$

Finally, and as emphasized by Hall (1997), the marginal rate of substitution may experience stochastic shifts—some of which might be permanent—both as a result of genuine shifts in individual tastes or composition effects derived from demographic forces. We can represent that feature by appending a preference shock ξ_t to our expression for the (log) marginal rate of substitution:

$$mrs_t = c_t + \varphi n_t + \xi_t. \tag{4}$$

Combining (2), (3), and (4), we obtain the following expression for hours:

$$n_t = -(1+\varphi)^{-1}(s_{c,t} + \mu_t^p + \mu_t^w + \xi_t).$$

Hence, we see that any permanent shifts in the labor and product market wedges or in preferences will result in a permanent shift in hours worked, for any given consumption share. To the extent that we are willing to allow for changes in any of those variables, there is no obvious reason (other than modeling convenience) why *all* those changes should be stationary. The presence of price and wage stickiness, with their consequent temporary deviations in markups from their desired levels, appears perhaps as the only "natural" source of stationarity in the fluctuations in hours in this context. Some changes in desired markups, on the other hand, are likely to be permanent (e.g., those resulting from changes in the regulatory environment). This property may also characterize changes in taxes that may affect the labor market wedge.[11] Needless to say, many shifts in preferences or in the share of consumption in GDP are likely to be permanent in nature.

Given the above considerations, the apparent nonstationarity in hours and its decoupling from the consumption share should come as no surprise, despite its central role in common versions of business cycle models.[12]

[11] Thus, Prescott (2004) argues that differences in labor income taxes can account for the gap in hours worked between the United States and Europe. Mulligan (2002) provides evidence of large and persistent changes in the labor market wedge.

[12] See Francis and Ramey (2002) for a similar argument using a fully specified growth model. In their model, permanent changes in government purchases, tax rates, or a preference parameter are shown to have a permanent effect on steady-state hours.

2.5 The Two Margins of Hours Variation: A Look at the International Evidence

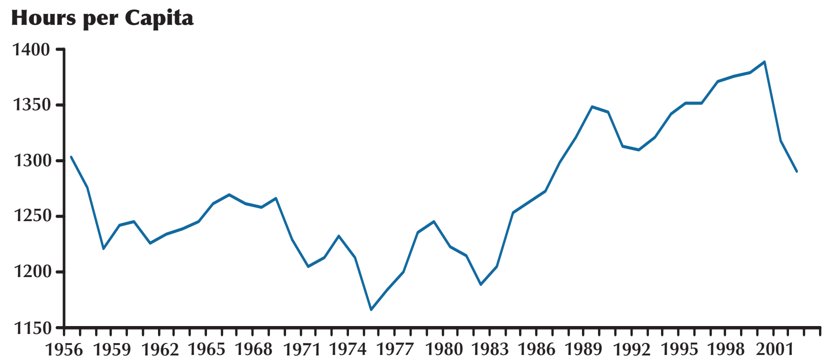
It is not the purpose of the present paper to uncover and even less so to quantify the relative importance of the different likely sources of nostationarity in hours. Yet, a look at the two margins of variation in hours, namely, hours per worker and the employment rate, suggests that a simple representative agent model will find it hard to account for the labor market dynamics behind the observed changes in hours. The low-frequency changes in both margins are substantial and display patterns that vary significantly across countries. These features are illustrated in Figures 3A through 3G, which show for each G7 country the evolution of hours worked per capita (more precisely, per person aged 16 to 65), average hours per worker, and the employment rate (i.e., the ratio of employment to working-age population). The data are drawn from the labor force statistics data set compiled by the OECD, with the sample periods differing somewhat across countries.

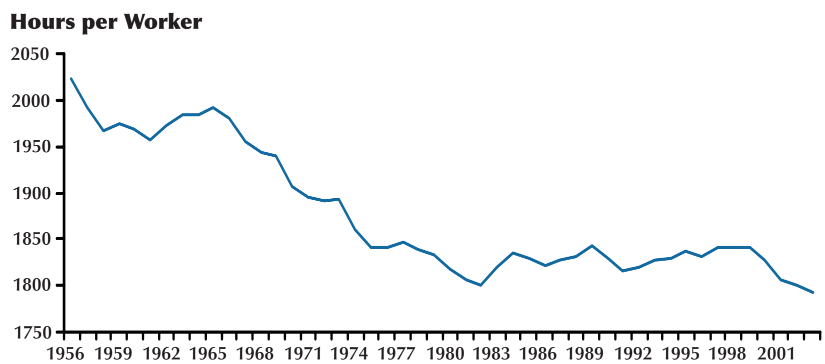
As a cursory look at the figures makes clear, the evidence constitutes an embarrassment of riches. Thus, for instance, the United States and Canada display an upward trend in hours per capita in the late part of the sample, which is the result of a continuous increase in the employment rate, combined with a flat pattern for hours per worker. In the early part of the sample, on the other hand, the relatively stable path for hours per capita hides very different (mutually offsetting) trends in hours per worker (downward) and the employment rate (upward).

The remaining G7 countries have experienced instead a secular decline in hours per capita. But, again, the underlying composition differs across countries. Thus, in the United Kingdom and Italy, it is largely the result of a decline in hours per worker, combined with a relatively stable employment rate. In Japan, hours per capita decline in spite of a persistent upward trend in the employment rate. In France and Germany, on the other hand, the downward trend in the employment rate only reinforces that of hours per worker.

Figure 3A

Labor Force Statistics: United States





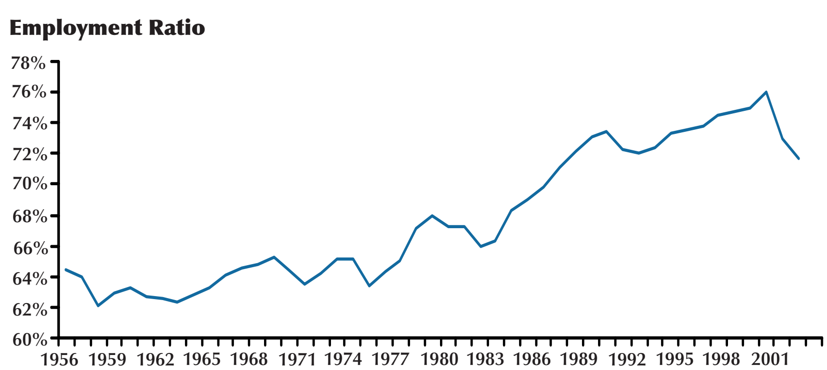


Figure 3B

Labor Force Statistics: Canada

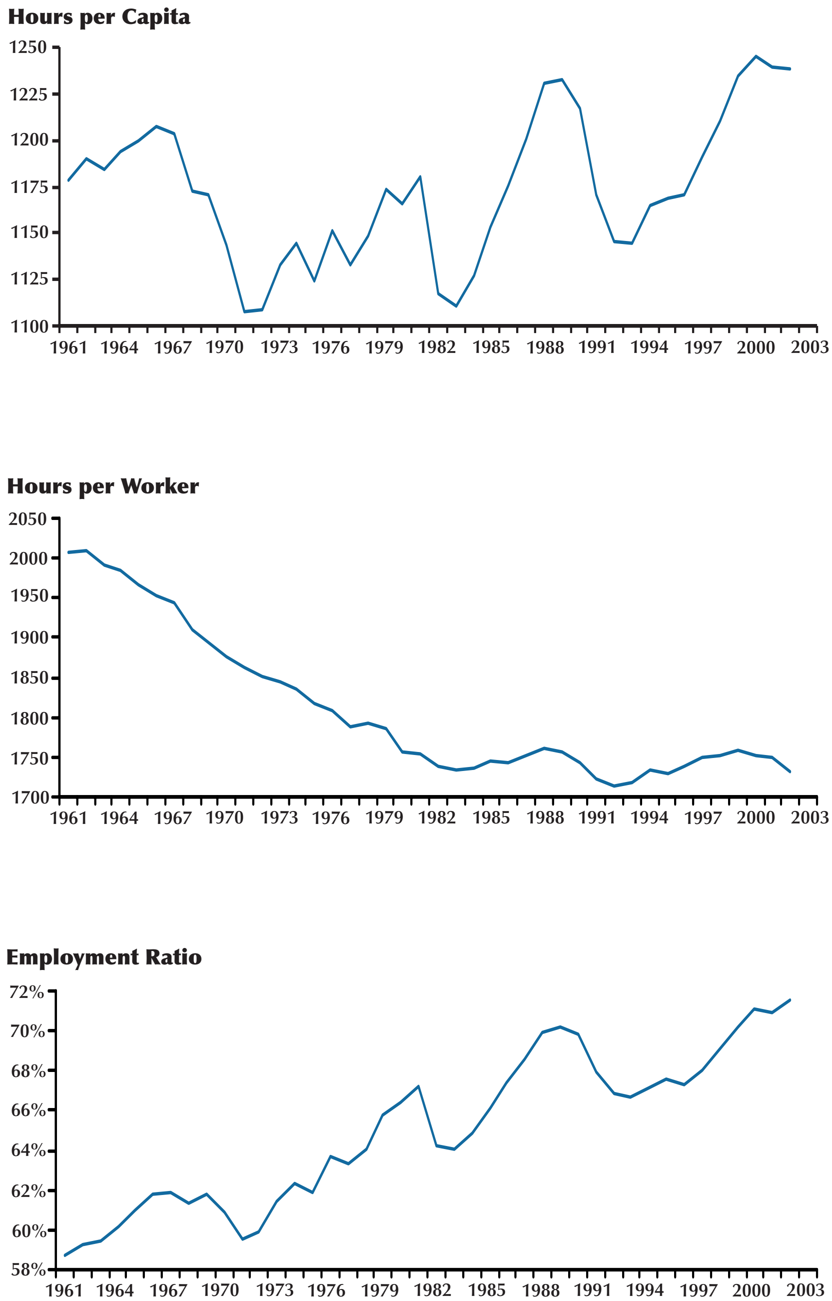


Figure 3C

Labor Force Statistics: Japan

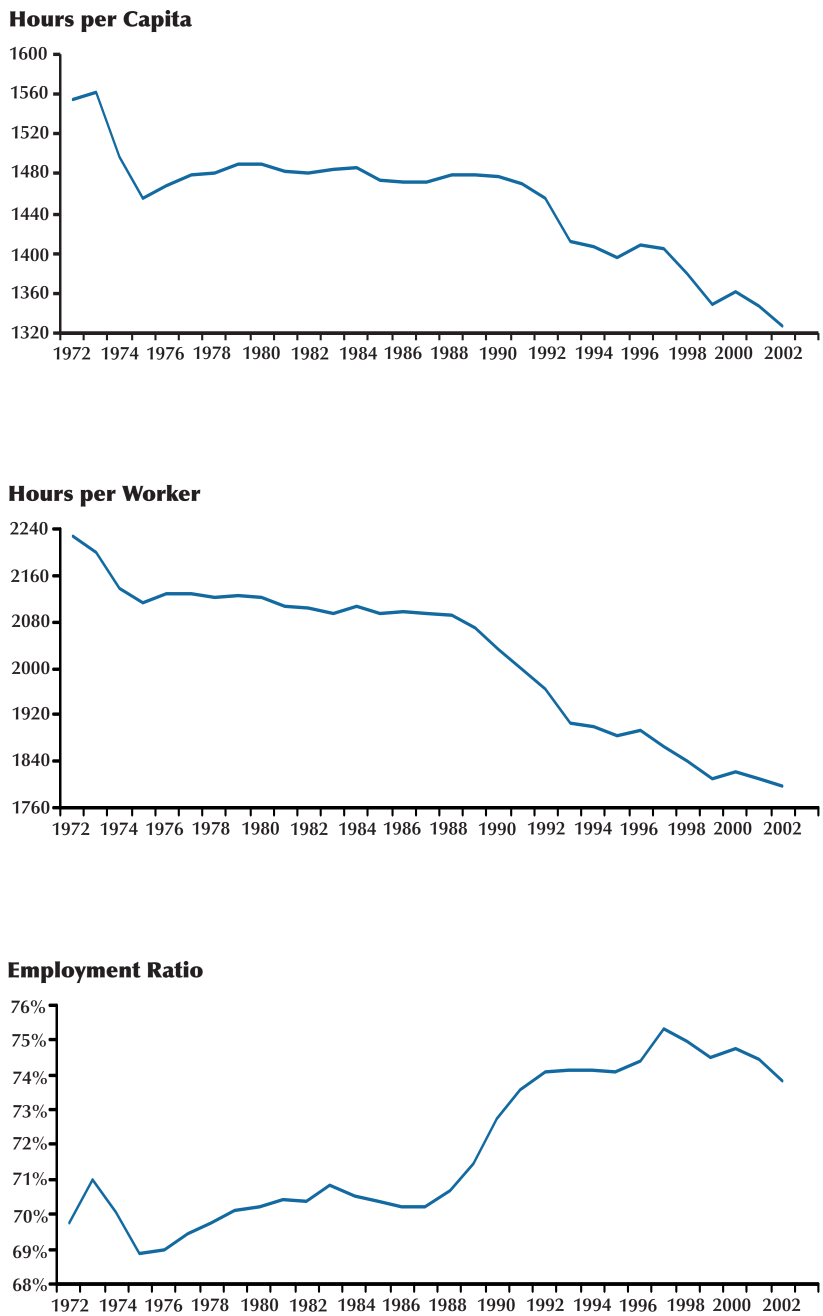


Figure 3D

Labor Force Statistics: United Kingdom

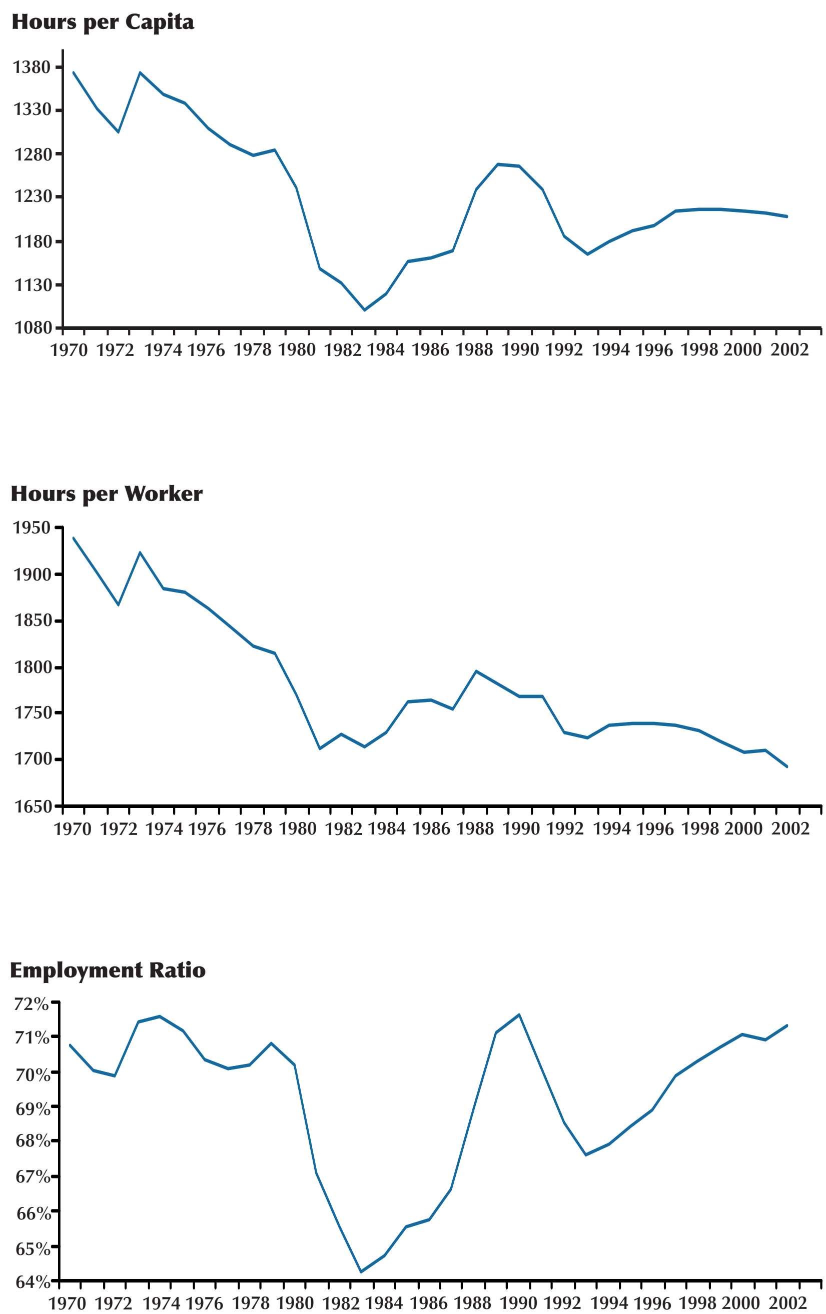
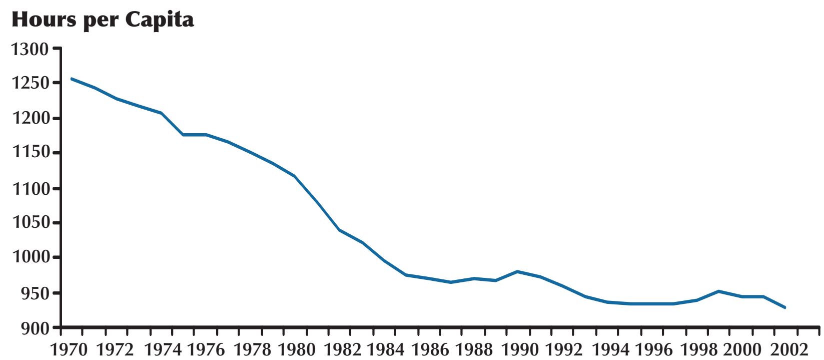
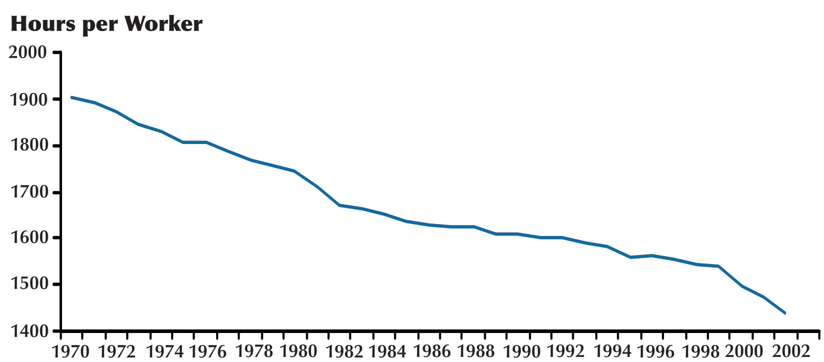


Figure 3E

Labor Force Statistics: France





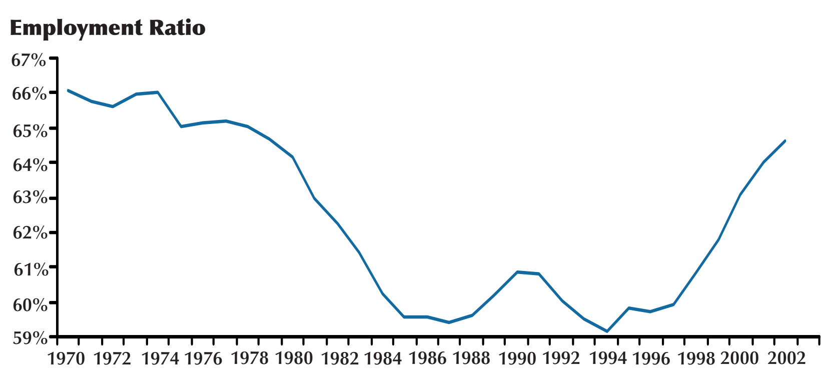


Figure 3F

Labor Force Statistics: Germany

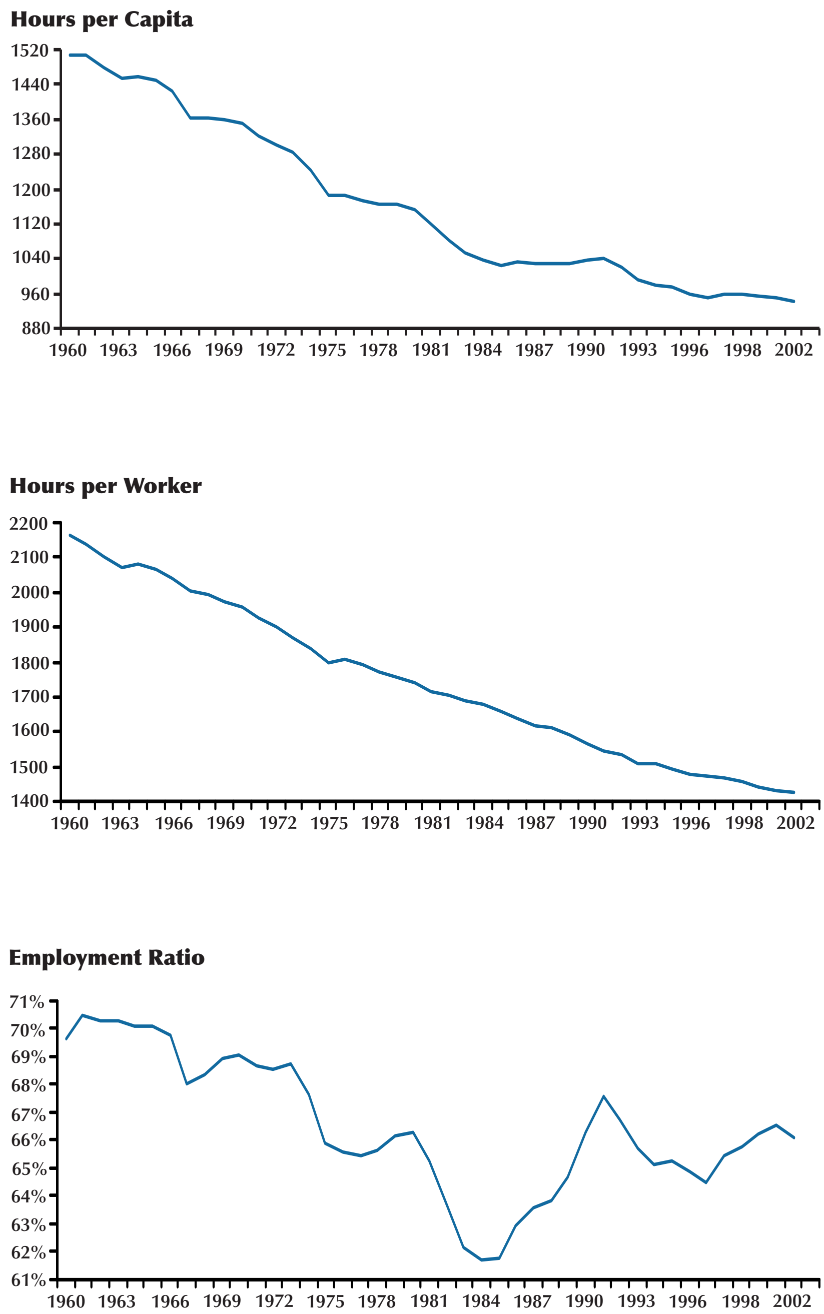
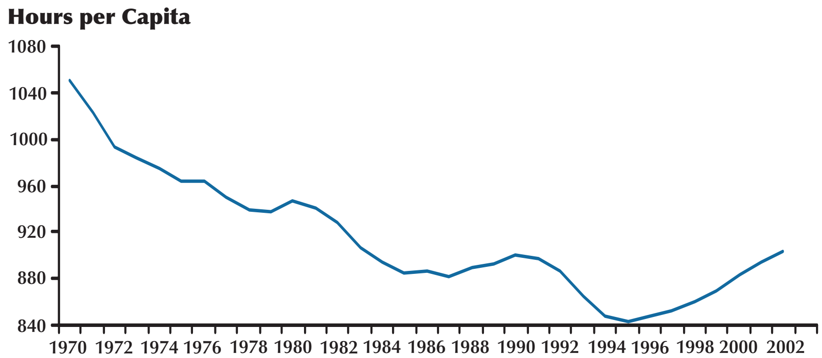
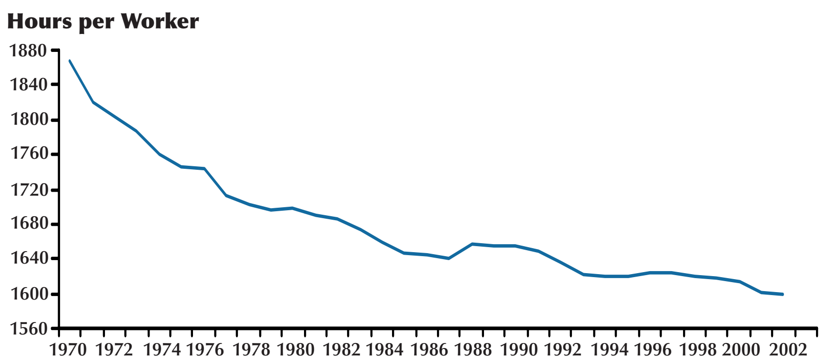
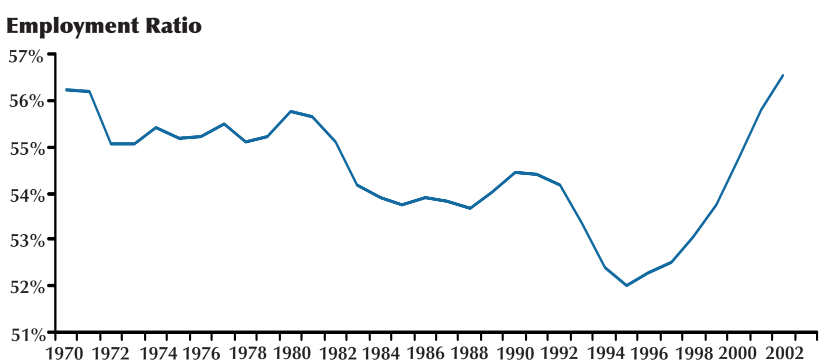


Figure 3G

Labor Force Statistics: Italy







2.6 Some Implications

While a number of papers have explored the consequences of introducing both margins of hours variations into a dynamic business cycle framework, I am not aware of any attempt to enrich those models with features that could help explain the low-frequency changes in hours per worker and the employment rate highlighted here previously.[13] To the extent that such low-frequency variations are due to factors orthogonal to the business cycle phenomena that those models seek to explain, abstracting from those features may be the right strategy when developing business cycle models. In particular, the secular decline in hours per worker observed in most countries would seem to fit that description. The same may be true for the observed long-run trend in the employment rate in the United States, Canada, and Japan. In contrast, a look at the low-frequency variation of the latter variable in the continental European countries suggests a stronger connection to cyclical episodes.

Most importantly, and putting aside the implications for the theoretical modeling of the business cycle, the existence of such low-frequency movements in hours is likely to impinge on *empirical* analyses and characterizations of economic fluctuations and their sources. The controversy, described later, regarding the appropriate transformation of hours in the recent empirical literature on the effects of technology shocks is a case in point. The evidence presented here contrasts starkly with the assumption of stationarity in hours found in many empirical applications and raises some doubts about any empirical findings that ignore the existence of those low-frequency movements, while hinging on the assumption of stationary hours. Furthermore, as discussed later, the existence of a significant nonstationary component in hours per worker implies that employment-based measures of productivity will be ridden with a nonstationary measurement error, which may question the validity of analyses that made use of those measures. That was the case, due to lack of data availability, for the international evidence on the effects of technology shocks reported in Galí (1999). In the next section I revisit that evidence using data on hours and hours-based productivity measures for the G7 countries.

3 INTERNATIONAL EVIDENCE ON THE EFFECTS OF TECHNOLOGY SHOCKS

3.1 Background

The dynamic effects of technology shocks and their role as a source of economic fluctuations have been the focus of growing interest among macroeconomists.[14] The bulk of the evidence generated by that literature has pointed to effects of technology very much at odds with the predictions of the standard RBC model. In particular, when technology shocks are identified in a structural vector autoregression (VAR) as the source of permanent shifts in labor productivity, they generate a negative comovement between hours and output (see, e.g., Galí, 1999, and Francis and Ramey, 2002). Furthermore, and not surprisingly in light of the previous result, the implied estimated contribution of technology shocks to the variability of output and hours at business cycle frequencies tends to be very small, an observation that contrasts with the central role assigned to those shocks as a source of business cycles in the RBC literature.[15]

In a recent paper, Christiano, Eichenbaum, and Vigfusson (2003; CEV hereafter) have pointed out that the first finding (i.e., the negative response of hours to a positive technology shock) may not be robust to using an alternative VAR specification that includes the level of (log) hours, as opposed to detrended hours or its first difference as a labor input measure. More specifically, when CEV reestimate the VAR using the *level* of (log)

[13] A possible exception, albeit in the context of a stylized model of wage-setting by unions/insiders with no explicit variation at the intensive margin, is found in Blanchard and Summers (1986).

[14] For a recent survey of that literature, see Galí and Rabanal (2004).

[15] Similar results are obtained in Basu, Fernald, and Kimball (forthcoming) using a growth-accounting approach to construct a measure of technological change. See also the references in the Galí and Rabanal (2004) survey.

hours, they find that a positive technology shock drives hours up, not down, and generates the positive comovement of hours and output that is the hallmark of the business cycle.[16]

A recent paper by Fernald (2004) provides a convincing explanation for the discrepancy of the estimates of the effects of technology shocks when hours are introduced in levels. In particular he shows that the presence of a low-frequency correlation between labor productivity growth and per capita hours, while unrelated to cyclical phenomena, distorts significantly the estimates of the short-run responses. Fernald illustrates that point most forcefully by reestimating the structural VAR in its levels specification (as in CEV), while allowing for two (statistically significant) trend breaks in labor productivity (in the mid-1970s and mid-1990s, respectively) and showing that hours decline in response to a positive technology shock in the resulting estimates, in a way consistent with the findings obtained when the difference specification is used.

A different strategy is pursued by Francis and Ramey (2004). Those authors construct an alternative time series for hours, with an annual frequency and spanning the entire 20th century. The Francis-Ramey series normalizes total hours worked in the business sector using a population measure that excludes not only the population aged 16 and less (as in CEV), but also the population older than 65, as well as that enrolled in school or employed by the government. The resulting series for hours per capita is largely devoid of any of the nonstationary features that were apparent in the original series (as discussed in Section 2). Most interestingly, though, when Francis and Ramey reestimate the effects of technology shocks using the (log) level of the new series, they recover the findings associated with the detrended or first-differenced specification, i.e., a positive technology shock induces a short-run decline in hours and a negative comovement between the latter variable and output.

[16] Yet, it is important to stress that their estimates of the contribution of technology shocks to the variance of output and hours is rather small, consistent with the findings in Galí (1999) and others.

3.2 New Evidence for the G7 Countries

Next I present some further evidence on the effects of technology shocks using data for the G7 countries. Motivated by the evidence and discussion in Section 2, I treat (log) hours (as well as labor productivity) as a difference-stationary series. Accordingly, I estimate for each G7 country the structural model

$$\begin{bmatrix} \Delta x_t \\ \Delta n_t \end{bmatrix} = \begin{bmatrix} C^{11}(L) & C^{12}(L) \\ C^{21}(L) & C^{22}(L) \end{bmatrix} \begin{bmatrix} \varepsilon_t^z \\ \varepsilon_t^d \end{bmatrix} \equiv C(L)\varepsilon_t \,, \tag{5}$$

where ε_t^z and ε_t^d are serially uncorrelated, mutually orthogonal structural disturbances whose variance is normalized to unity. The polynomial $|C(z)|$ is assumed to have all its roots outside the unit circle. Estimates of the distributed lag polynomials $C^{ij}(L)$ are obtained by a suitable transformation of the estimated reduced form VAR for $[\Delta x_t, \Delta n_t]$ after imposing the long-run identifying restriction $C^{12}(1) = 0$.[17] That restriction effectively defines $\{\varepsilon_t^z\}$ and $\{\varepsilon_t^d\}$ as shocks *with* and *without* a *permanent* effect on labor productivity, respectively. On the basis of some of the steady-state restrictions shared by a broad range of macro models, I interpret permanent shocks to productivity $\{\varepsilon_t^z\}$ as *technology shocks*. On the other hand, transitory shocks $\{\varepsilon_t^d\}$ can potentially capture a variety of driving forces behind output and labor input fluctuations that would not be expected to have permanent effects on labor productivity. I refer the reader to Galí (1999) and Francis and Ramey (2002) for a detailed discussion.

A number of recent papers have provided related evidence using data for countries other than the United States, as surveyed in Galí and Rabanal (2004). Here I revisit some earlier evidence using international data, which the findings of Section 2 might call into question. Thus, in Galí (1999, 2004) and Francis, Owyang, and Theodorou (2004), a similar structural VAR-based approach is applied to G7 countries' quarterly data. In that exercise, Galí (1999) used both first-differenced and detrended (log) employment as a measure of labor input, obtaining a negative response of employment to a positive technology shock in all countries, with the exception of Japan.

[17] See Blanchard and Quah (1989) and Galí (1999) for details.

Similar qualitative results for the euro area as a whole are found in Galí (2004), where I used the quarterly data set for the euro area recently constructed by Fagan, Henry, and Mestre (2001), also using employment as a measure of labor input.[18] Francis, Owyang, and Theodorou (2004) used first-differenced (log) employment, together with other macro variables, focusing on the potential role of monetary policy factors in shaping the response to technology shocks.

In light of the evidence and discussion found in Section 2, the use of employment as a measure of labor input raises a potential problem. First, and to the extent that hours per worker vary over time, employment-based measures of labor productivity will be ridden with error. Most importantly, however, the presence of a possible nonstationary component in hours per worker undermines the theoretical basis for the identification strategy used in that work. In particular, it implies that factors other than technology may generate permanent shifts in measures of output per worker—and hence may be incorrectly labeled as technology shocks—even if they do not have any long-run effect on "true" labor productivity (i.e., output per hour). The availability of (relatively long) homogeneous time series for hours worked in a number of OECD countries allows one to overcome this problem. Next I describe the evidence on the effects of technology shocks obtained by applying the structural VAR framework previously noted to hours-based measures of productivity and labor input for the G7 countries.

The top graph in the top panel of Figures 4A through 4G displays, for each G7 country, the growth rates of GDP and hours over the period 1970-2002. As it is clear to the eye, both series display a strong positive comovement, which can be viewed as one of the defining features of the business cycle. The corresponding unconditional correlations, reported in the third column of Table 3, range from 0.30 (Italy) to 0.84 (United States). The second graph shows the components of GDP and hours growth associated with technology shocks, while the third graph shows the component driven by other (nontechnology) sources. With the exception of the United Kingdom and Japan, the comovement of hours and GDP growth generated by technology shocks is much smaller than that observed in the original data. As shown in Table 3, that *conditional* correlation is even negative in some cases (United States, Italy, and France) and very small in others (Germany and Canada). The first two columns of Table 3, which show the estimated contribution of technology shocks to the variance of GDP and hours growth, suggest a limited role of technology shocks as a source of fluctuations in the growth rates of either GDP or hours. In particular, technology shocks account for more than 50 percent of the variance of hours growth only in Italy and the United Kingdom, though as discussed below, this can hardly be attributed to the mechanisms underlying RBC model.

A look at the estimated impulse responses of labor productivity, GDP, and hours to a positive technology shock (i.e., one that raises productivity) helps clarify the previous patterns. Those impulse responses (together with a ± 2 standard errors confidence interval) are shown in the bottom panel of Figures 4A-4G. The sign of the point estimate for the impact responses of GDP and hours is also shown in Table 3. For all countries but Japan, a positive technology shock generates either a negligible response of hours or a short-run decline in that variable, in contrast with the predictions of the standard RBC model. With the exception of the United Kingdom, the corresponding response of output to the same shock is positive. It is thus not surprising that for the remaining countries (United States, Canada, Germany, Italy, and France) technology shocks do not make the large contribution to the variance of *both* GDP and hours growth that proponents of the RBC paradigm would have led us to expect. In fact, a look at Table 3 suggests that the pattern of conditional second moments and impulse responses that one would associate with an RBC model that were capable of accounting for the bulk of GDP and hours fluctuations can only be found for Japan. Why the latter country (as well as the United Kingdom) displays a pattern different from the rest is beyond the scope of this paper.

[18] In particular, technology shocks are found to account for only 5 percent and 9 percent of the variance of the business cycle component of euro area employment and output, respectively, with the corresponding correlation between their technology-driven components being –0.67.

Figure 4A

Sources of Fluctuations: United States

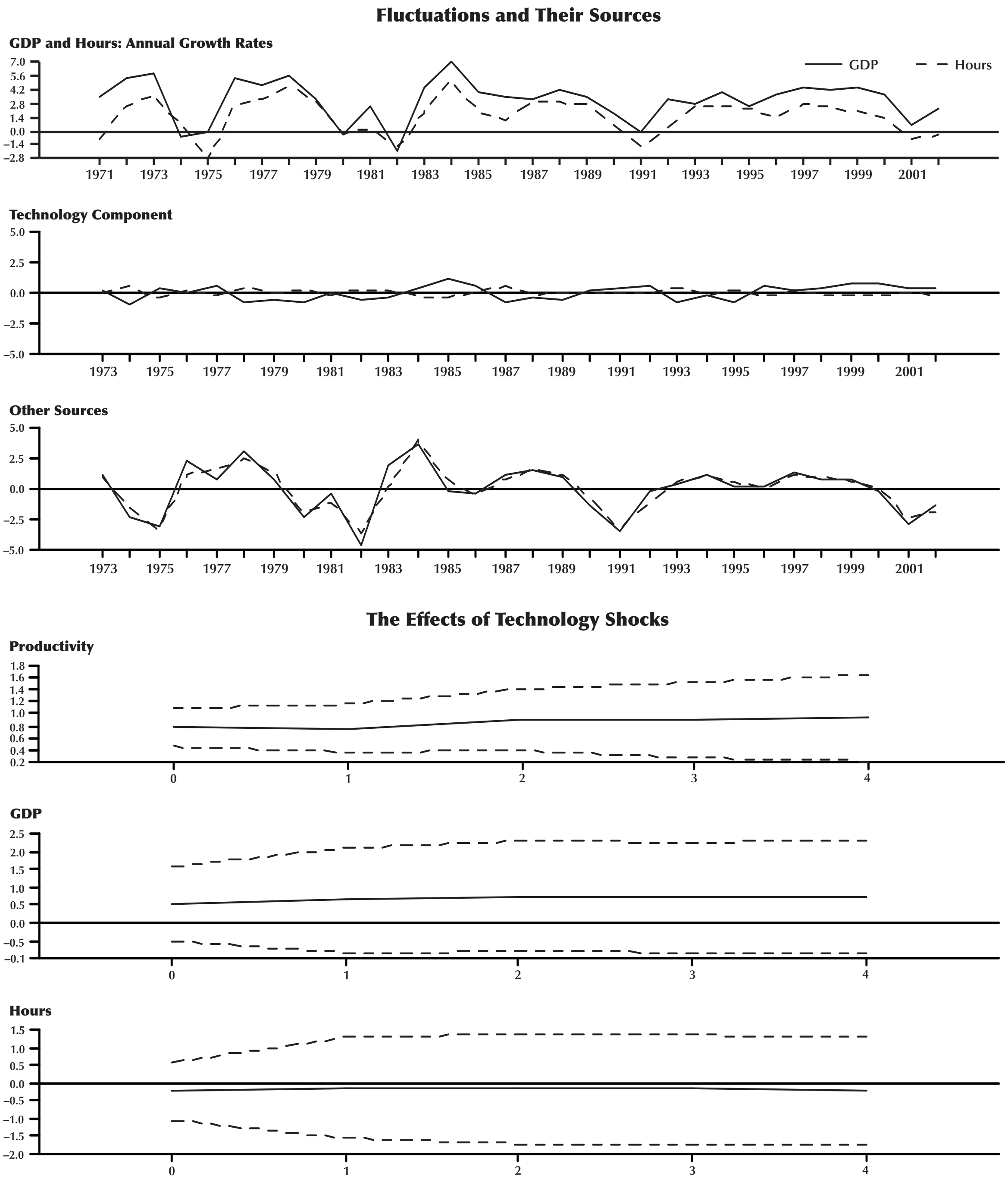


Figure 4B

Sources of Fluctuations: Canada

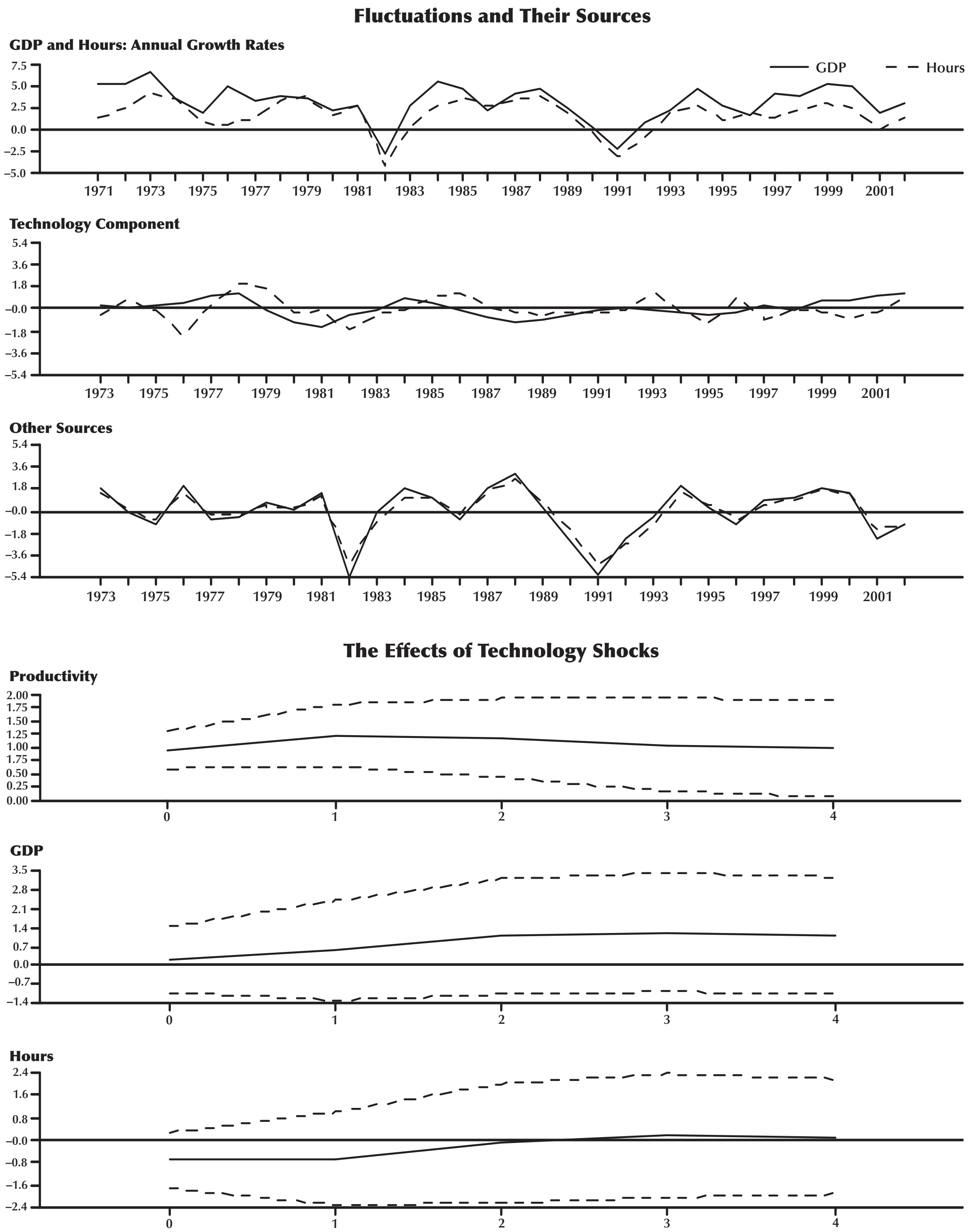


Figure 4C

Sources of Fluctuations: Japan

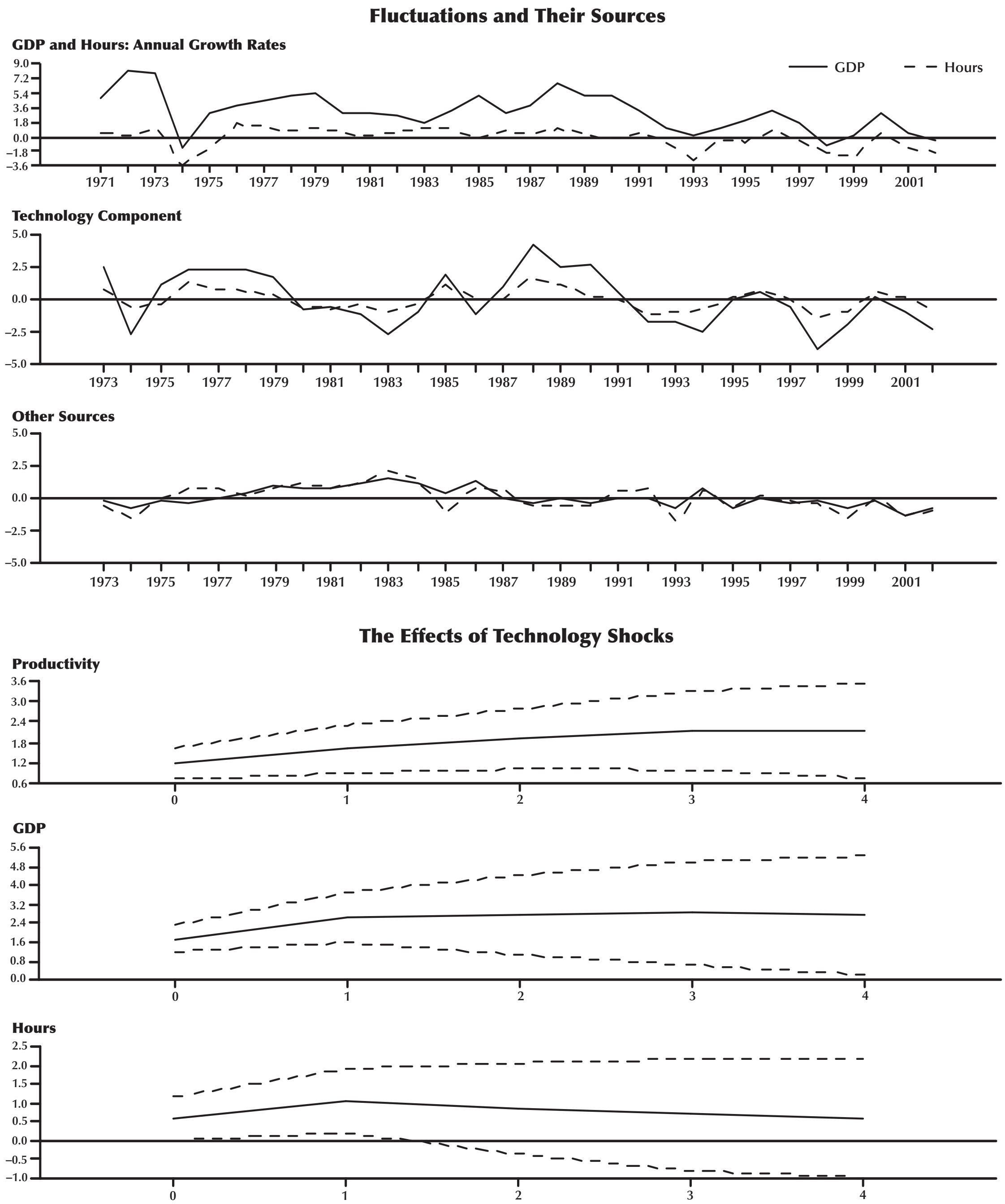


Figure 4D

Sources of Fluctuations: United Kingdom

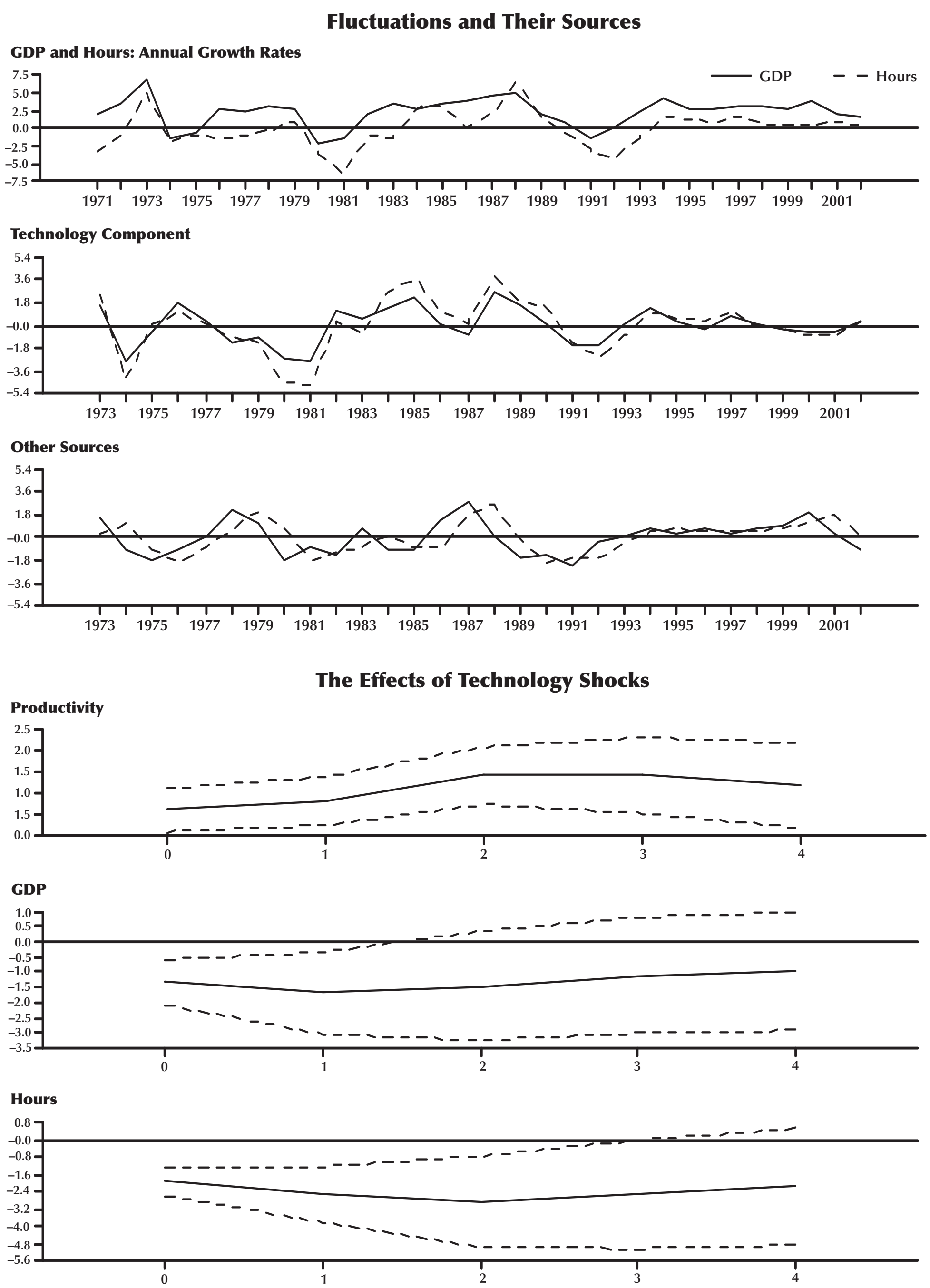


Figure 4E

Sources of Fluctuations: Germany

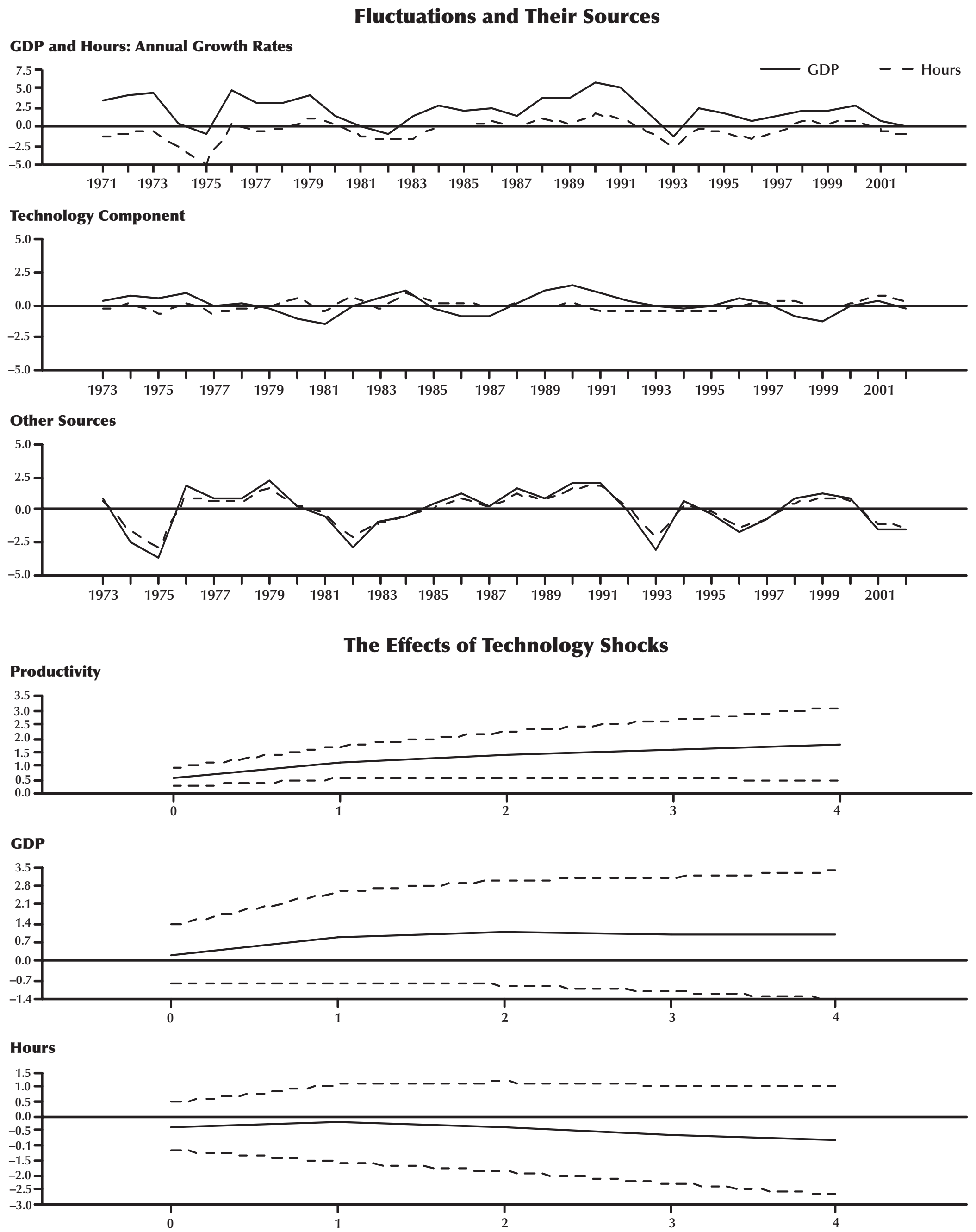


Figure 4F

Sources of Fluctuations: France

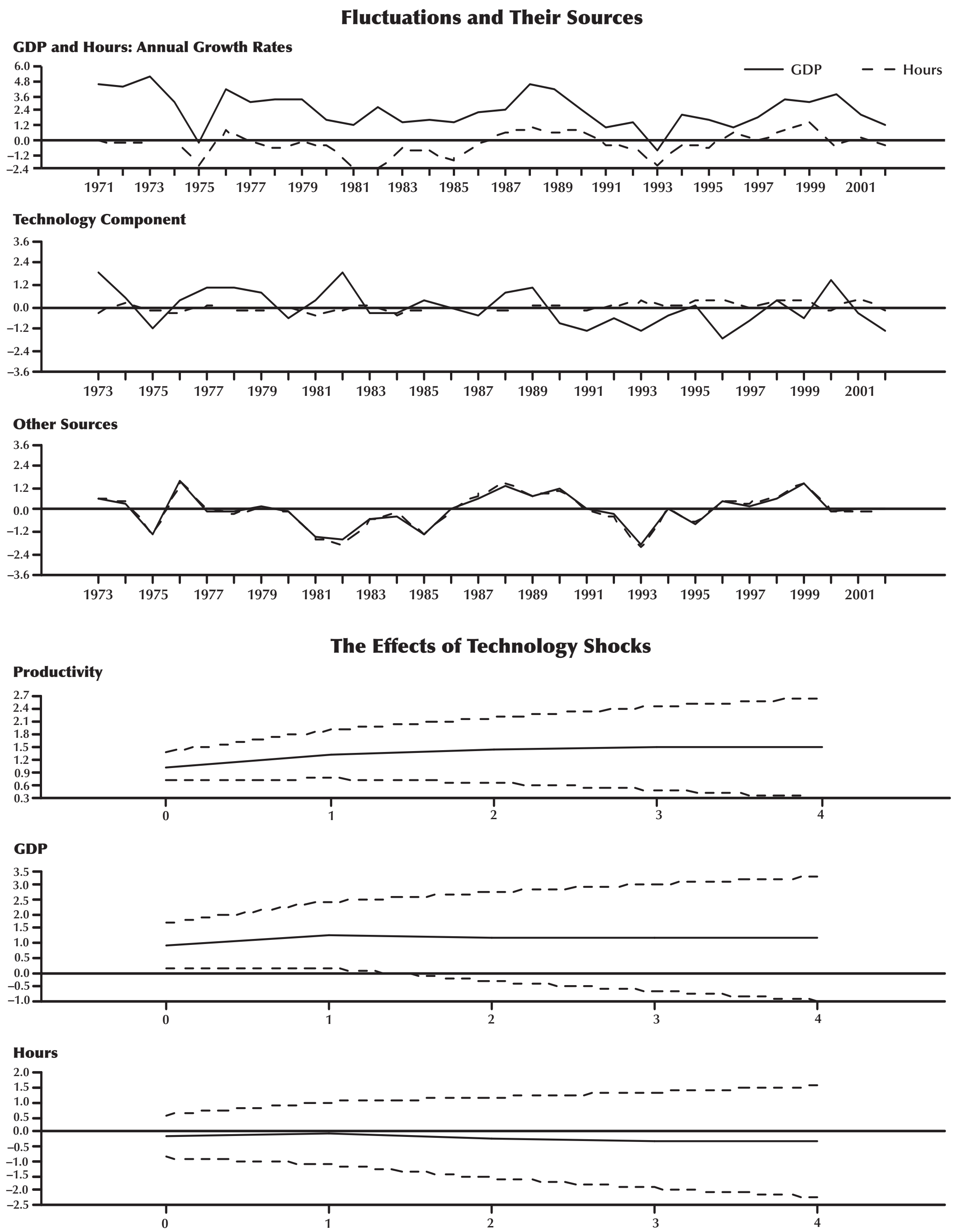


Figure 4G

Sources of Fluctuations: Italy

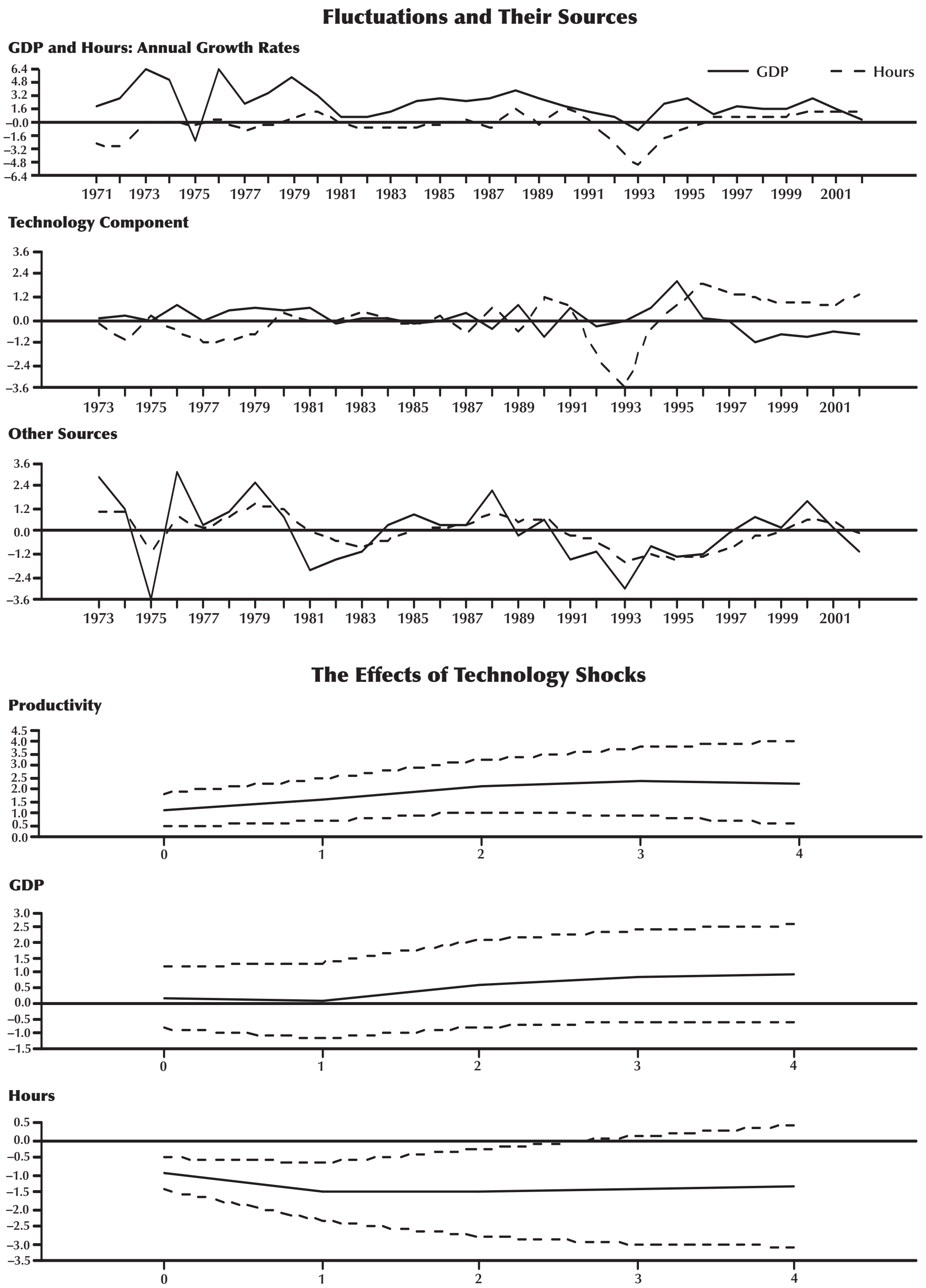


Table 3

Technology Shocks and Business Cycles: International Evidence

	Variance share		Correlation (Δn,Δy)		Impact response	
	var(Δy)	var (Δn)	Unconditional	Conditional	Hours	GDP
United States	0.09	0.02	0.84	–0.79	–	+
Canada	0.12	0.24	0.83	0.24	–	+
Japan	0.88	0.40	0.70	0.89	+	+
United Kingdom	0.54	0.74	0.76	0.93	–	–
Germany	0.17	0.11	0.73	0.07	–	+
Italy	0.13	0.65	0.30	–0.26	–	+
France	0.54	0.07	0.56	–0.32	–	+

4 SUMMARY AND CONCLUSIONS

The present paper has revisited the empirical and theoretical grounds for a property found in most dynamic macroeconomic models—namely, that of stationarity of hours worked per capita. From a theoretical viewpoint, I have argued that stationarity of hours per capita is not a necessary condition for a macro model to generate a balanced-growth path. One can think of many factors that could lead to nonstationary hours, including permanent shifts in government purchases or in labor and product market wedges, as well as permanent preference shifts. In fact, it is hard to imagine why some of those factors would remain unchanged or display only transitory fluctuations.

From an empirical perspective, I have shown that the evidence for most industrialized economies appears to be at odds with the equilibrium relationships that are at the root of the property of stationarity in hours per capita. In fact, that evidence suggests that both margins of labor utilization (i.e., hours per worker and the employment rate) display some nonstationarity features in most countries.

Finally, I have revisited the international evidence on the effects of technology shocks using a measure of output per hour for the G7 countries. That measure overcomes the potential identification problems that may have resulted from using an employment-based measure of productivity, given the evidence of significant nonstationarity in hours per worker. The findings from that exercise suggest that, possibly with the exception of Japan, the evidence on the effects of technology shocks shows major discrepancies with the predictions of standard real business cycle models.

REFERENCES

Basu, Susanto; Fernald, John and Kimball, Miles. "Are Technology Improvements Contractionary?" *American Economic Review* (forthcoming).

Baxter, Marianne, and King, Robert G. "Measuring Business Cycles: Approximate Band-Pass Filters for Economic Time Series." *Review of Economics and Statistics*, 1999, *81*(4), pp. 575-93.

Blanchard, Olivier J. and Quah, Danny. "The Dynamic Effects of Aggregate Demand and Supply Disturbances." *American Economic Review*, 1989, *79*(4), pp. 654-73.

Blanchard, Olivier J. and Summers, Lawrence. "Hysteresis and the European Unemployment Problem," in Stanley Fischer, ed.; *NBER Macroeconomics Annual 1986*. Cambridge, MA: MIT Press, 1986, pp. 15-77.

Christiano, Larry; Eichenbaum, Martin and Evans, Charles. "Nominal Rigidities and the Dynamic Effects of a Shock to a Monetary Policy." Working Paper No. 0107, Federal Reserve Bank of Cleveland (forthcoming in *Journal of Political Economy*).

Christiano, Lawrence; Eichenbaum, Martin and Vigfusson, Robert. "What Happens after a Technology Shock?" NBER Working Paper No. 9819, National Bureau of Economic Research, 2003.

Cooley, Thomas F. and Prescott, Edward C. "Economic Growth and Business Cycles," in T.F. Cooley ed., *Frontiers of Business Cycle Research*. Princeton, NJ: Princeton University Press, 1995.

Fagan, Gabriel; Henry, Jerome and Mestre, Ricardo. "An Area-Wide Model (AWM) for the Euro Area." ECB Working Paper No. 42, European Central Bank, 2001.

Fernald, John. "Trend Breaks, Long Run Restrictions, and the Contractionary Effects of a Technology Shock." Unpublished manuscript, Federal Reserve Bank of Chicago, 2004.

Francis, Neville; Owyang, Michael T. and Theodorou, Athena T. "What Explains the Varying Monetary Response to Technology Shocks in G-7 Countries?" Working Paper 2004-002C, Federal Reserve Bank of St. Louis, 2004.

Francis, Neville and Ramey, Valerie. "Is the Technology-Driven Real Business Cycle Hypothesis Dead? Shocks and Aggregate Fluctuations Revisited." NBER Working Paper No. 8726, National Bureau of Economic Research, 2002 (forthcoming in *Journal of Monetary Economics*).

Francis, Neville and Ramey, Valerie. "A New Measure of Hours per Capita with Implications for the Technology-Hours Debate." Unpublished manuscript, 2004.

Galí, Jordi. "Technology, Employment, and the Business Cycle: Do Technology Shocks Explain Aggregate Fluctuations?" *American Economic Review*, 1999, *89*(1), pp. 249-71.

Galí, Jordi. "On the Role of Technology Shocks as a Source of Business Cycles: Some New Evidence." *Journal of the European Economic Association*, 2004, *2*(2-3), pp. 372-80.

Galí, Jordi; Gertler, Mark and López-Salido, David. "Markups, Gaps and the Welfare costs of Business Fluctuations." Unpublished manuscript, 2003.

Galí, Jordi and Rabanal, Pau. "Technology Shocks and Aggregate Fluctuations: How Well Does the Real Business Cycle Model Fit Postwar U.S. Data?" NBER Working Paper No. 10636, National Bureau of Economic Research, 2004 (forthcoming in *NBER Macroeconomics Annual*).

Hall, Robert E. "Macroeconomic Fluctuations and the Allocation of Time." *Journal of Labor Economics*, 1997, *15*(1, Part 2), pp. S223-50.

Hansen, Gary D. "Indivisible Labor and the Business Cycle." *Journal of Monetary Economics*, 1985, *16*, pp. 309-27.

King, Robert G.; Plosser, Charles and Rebelo, Sergio T. "Production, Growth and Business Cycles: I. The Basic Neoclassical Model." *Journal of Monetary Economics*, 1988, *21*, pp. 195-232.

Kydland, Finn E. and Prescott, Edward C. "Time to Build and Aggregate Fluctuations." *Econometrica*, 1982, *50*, pp. 1345-70.

Mulligan, Casey. "A Century of Labor-Leisure Distortions." NBER Working Paper No. 8774, National Bureau of Economic Research, 2002.

Prescott, Edward C. "Theory Ahead of Business Cycle Measurement." Federal Reserve Bank of Minneapolis *Quarterly Review*, 1986, *10*, pp. 9-22.

Prescott, Edward C. "Why Do Americans Work So Much More Than Europeans?" Federal Reserve Bank of Minneapolis *Quarterly Review*, July 2004, *28*(1), pp. 2-13.

Rogerson, Richard. "Indivisible Labor, Lotteries and Equilibrium." *Journal of Monetary Economics*, 1988, *21*, pp. 3-16.

Sbordone, Argia. "An Optimizing Model of U.S. Wage and Price Dynamics." Unpublished manuscript, 2000.

APPENDIX

HOURS WORKED AND THE CONSUMPTION SHARE IN A TWO-SECTOR MODEL

The first-order condition of the household problem is

$$w_t - p_t^c = c_t + \varphi n_t,$$

where p_t^c denotes the price of consumption goods.

The first-order condition for a firm producing consumption goods is

$$w_t - p_t^c = y_t^c - n_t^c,$$

whereas the one for firms in the investment goods sector is

$$w_t - p_t^i = y_t^i - n_t^i.$$

A weighted average of the previous conditions implies

$$w_t - p_t = y_t - n_t,$$

which, combined with the expression derived above, yields

$$(p_t + y_t) = (p_t^c + c_t) + (1 + \varphi)\, n_t.$$

Letting $s_{c,t} \equiv (p_t^c + c_t) - (p_t + y_t)$ denote the (log) ratio of nominal consumption to nominal output and rearranging terms yields equation (1) in the text.

Commentary

Christopher A. Sims

THE CONTEXT

Galí's paper (2005) is difficult to understand unless one places it in the context of the series of papers, set off by Galí's 1999 paper, which investigates how much of the business cycle is accounted for by "technology shocks." The 1999 paper found that little of the business cycle was accounted for by technology shocks and that technology shocks caused productivity and labor input to move in opposite directions, contrary to the pattern of most business cycle fluctuations. Subsequently, in the paper Galí labels CEV, Christiano, Eichenbaum, and Vigfusson (2003) showed that in a two-variable or multivariable vector autoregression (VAR) identified by long-run restrictions, one could obtain a quite different result if one used data in levels rather than first differences. Chari, Kehoe, and McGrattan (2004) have attacked structural VARs in general, apparently motivated by their disbelief in the original Galí result. Galí and Rabanal have surveyed this literature and connected it to related literature. The Galí and Rabanal paper (2004; henceforth GR) is the place to start if one is interested in this literature, as it considers a wide variety of previous work and makes some nicely executed contributions of its own.

The conclusion in GR is still that technology shocks are not the main cause of business cycles. The most convincing evidence in GR is the results from the multivariate equilibrium model estimated in the paper and from the similar multivariate equilibrium models that have been fitted to U.S. and European data by Smets and Wouters (2003a,b). The Smets and Wouters models are validated by careful comparison of their statistical fit to that of Bayesian VARs. These models suggest a contribution of technology shocks of about 15 to 35 percent of business cycle variance, in contrast to the under-10 percent estimates in Galí's paper for this conference (2005) and his original (1999) paper. GR show that the estimated contribution of neutral technology shocks does not rise to the high levels suggested by the early real business cycle (RBC) literature unless all of a long list of frictional mechanisms are shut down. They do not compute posterior odds ratios, and indeed the error bands they display for impulse responses are so narrow that it seems likely that their model does not compete with Bayesian VARs in fit. Still, their results roughly match those of the Smets and Wouters models, in which we know that shutting down these frictions seriously impairs the fit.

Though one can read GR as confirming the original Galí paper's conclusions, GR does represent some movement away from the original paper's conclusions and a major step away from its methodology. GR acknowledge that there can be technology shocks that do not have a long-run impact on productivity, and indeed that such shocks emerge in estimated dynamic stochastic general equilibrium (DSGE) models as important and as inducing positive comovement between labor input and productivity. The possibility of drifting productivity due to nontechnology shocks

Christopher A. Sims is a professor of economics at Princeton University and a visiting scholar at the Federal Reserve Banks of New York and Atlanta.

Federal Reserve Bank of St. Louis *Review*, July/August 2005, *87*(4), pp. 487-92.

is also acknowledged, with capital taxes discussed explicitly; drifts in time preference, which they don't discuss explicitly, might in principle be equally important. And as we have already noted, the estimated contribution of technology shocks to variance is in GR much greater than in the earlier Galí paper and in this paper.

So the substantive conclusion from GR, as I read it, is that technology shocks are very likely important enough that Keynesians of the early 1970s would have found the results surprising, even if they are also very likely not nearly as important as suggested in the earliest RBC models. In fact, it is a bit ironic that the fitted DSGE models imply similar nontrivial but modest roles for monetary policy shocks and technology shocks. It is as if the data are telling us that extremist monetarists, Keynesians, and RBC calibrators are all wrong, but all have a piece of the truth.

The methodological conclusion from the GR paper, as I read it, is that models with one or two shocks and/or one or two variables should be set aside. The estimated DSGE models do not imply that such small models can cleanly separate two meaningful categories of disturbances to the economy. There is plenty of room for methodological improvement, however, and this could change substantive conclusions. The entire literature, including GR, struggles with issues of detrending, differencing or not differencing, Hodrick-Prescott (HP) filtering, etc. Even Smets and Wouters, who do the best job so far of integrating various sources of uncertainty into substantive conclusions, use ad hoc detrending methods and do not fully incorporate uncertainty about low-frequency behavior into their analysis.

The paper at hand, though it does not reach this conclusion explicitly, is in fact a confirmation of the points that bivariate models are inadequate and that uncertainty about low frequencies is central. It documents drifting behavior in consumption share and hours per worker in all the countries it studies. It provides potential explanations for these drifts that amount to postulating additional sources of disturbance, which implies that a larger model would be useful, or perhaps even necessary, to sort out sources of variation. And its results are inconclusive and variable across countries.

LONG- AND SHORT-RUN INFERENCE

From one perspective, the RBC innovation was to insist that we should integrate the theoretical frameworks in which we analyzed growth and business cycle fluctuations. It has always seemed to me paradoxical, therefore, that the convention in the RBC literature has been to filter low-frequency variation out of the data before proceeding to analyze business cycle variations. The practical reason for this is similar to that underlying the use of deseasonalized data: The low-frequency data are *extremely informative* about parameters of simple growth models, so that if we fit freely to all the data, the resulting model would be essentially determined by the very-low-frequency data, leaving a poor fit to the cyclical frequencies. Of course, this is not a necessary outcome in principle. It is because of unrealistic simplicity in model dynamics that they cannot at the same time match low-frequency and higher-frequency data.

With seasonality, the additional model complexity required to fit seasonal and cyclical variation simultaneously is arguably a poor trade-off, because seasonality involves phenomena—like weather and holidays—that bring in new parameters unrelated to our central interests. With growth, though, the additional model complexity required would essentially be just more flexible and realistic modeling of sources of inertia, not fundamentally new structural parameters. We could handle such models now.

The literature has persisted in using ad hoc detrending methods and ignoring the effects of the detrending on uncertainty. The best treatment of low-frequency variation so far is probably in the Smets and Wouters U.S. model, where they remove a common linear trend from most (logged) variables and account for uncertainty about the trend parameter in constructing their posterior distributions. The trend is not treated directly as a structural parameter, however, and sample means are extracted in advance with no accounting for uncertainty about them. In their model of the European Economic and Monetary Union, Smets and Wouters extracted a separate trend from each variable and did not account for uncertainty about the trends. Unsurprisingly, in light of these differences in treatment of low frequencies, the variance decomposition for the European data is quite different from that for the U.S. data, even though the impulse responses are qualitatively similar in the short run.

Galí's early paper and the one at hand use differenced log data. CEV pointed out that conclusions differed with variables in levels. GR agree, pointing also to work by other authors, that results are sensitive to whether data are differenced and to what kind of trend-removal is applied.

It appears to me that there are two reasons, beyond the fact cited above that standard models are not crafted to match low- and business cycle–frequencies simultaneously, for the persistence of detrending and differencing adhockery. One is the use of the HP filter in the early RBC literature and the strong tendency in the economics literature for the methodology of widely read papers to be imitated uncritically. The other is that, until recently, few economists understood Bayesian reasoning and hence most were inhibited by the formidable conceptual problems for inference about low frequencies in a frequentist approach. GR, though, do estimate a multivariate DSGE model using Bayesian methods. Bayesian methods can easily accommodate inference about means, trends, and orders of differencing. It is therefore disappointing that GR follow the rest of the literature in using a preliminary ad hoc detrending approach.

UNIT ROOTS, IDENTIFICATION

There is a fundamental problem, recognized years ago, with identification of VARs by means of long-run restrictions. Sums of coefficients in MA or AR operators are weakly identified—indeed identified only by means of lag length restrictions—unless the variables driving the operator are nonstationary. Without the nonstationarity, one can fix sums of coefficients arbitrarily while achieving fits arbitrarily close to that of the true model.

The sum of coefficients that drives identification in this paper's structural VAR exercise is a sum of coefficients on a lag operator that applies to a stationary variable—the nontechnology shocks in the MA form, the Δn variable in the AR form. It is therefore likely that the identification is fragile. Just to illustrate, suppose Δn is i.i.d. and that the true lag distribution on Δn in the equation defining the technology shock is a sequence of zeros. Suppose further that there is a rotation of the model that makes the lag distribution on Δn in that equation $\gamma(L) = L$, i.e., only lagged Δn enters the equation, with a coefficient of 1. How close could the mistaken rotation, in which $\gamma(1) \neq 0$, come to the fit of the true model while still satisfying the identifying assumption $\gamma(1) = 0$? The answer depends on lag length. If we fit a model in which $\gamma(L)$ is restricted to order 5, setting $\gamma(L) = 0.8L - 0.2L^2 - 0.2L^3 - 0.2L^4 - 0.2L^5$ gives a predictor with zero sum of coefficients that has $R^2 = 0.8$ with the false $\gamma(L) = 1$ predictor. With a lag length of n we can achieve an R^2 of $1 - 1/n$.

That identification in this setup depends on our treating lag length as known a priori is made very clear in the framework used by Shapiro and Watson (1988) and followed also in CEV. The estimation proceeds by using an equation of the form

$$\Delta f_t = \beta(L)\Delta f_{t-1} + \gamma(L)\Delta X_t + \varepsilon_t^z,$$

where f_t is productivity and X_t is a list of other, stationary variables. Current X is allowed in the equation, but is assumed to be possibly correlated with ε_t^z. β contains powers 1 to q of L and γ contains powers 0 to $q-1$. The solution to the simultaneity problem is to use lags 1 to q of Δf_t and lags 1 to q of X_t as instruments. It may be initially puzzling as to how this makes sense. The proposed instruments all seem to appear directly in the equation. But this is not quite true. X appears in the restricted equation only as ΔX. The instruments are the levels of X. It is true that all the lagged ΔX terms are exact linear combinations of the instruments, but current ΔX_t is not—quite. Identification conditions are formally satisfied because the best predictor of ΔX_t based on $q-1$ lags of ΔX_t is not as good as the best predictor based on q lagged levels of X_t. But as the little example above should make clear, the difference between these two predictors, and hence the firmness of the identification, quickly shrinks toward zero as q increases. Because in fact we do not know q, but adjust it by checking fit with various values of q—i.e., we estimate q—the zero sum of coefficients restriction is arguably no restriction at all.

Once we understand this knife-edge identification, it is unsurprising that apparently minor differences in specification can have major effects on results.

The SVAR work in this paper is conditioned on there being two, nonrepeated unit roots in the joint productivity and hours process. The data are consistent with this assumption. When I run for Italy and the United States a levels version of the reduced form VAR underlying this paper's panel of country SVARs, a weak Minnesota prior, which pulls gently toward the two-unit-roots hypothesis, manages to pull the point estimates to nearly exactly satisfying the two-unit-roots hypothesis.

But when I run that reduced form without a prior, the sums of coefficients matrices emerge as follows:

	Italy		United States	
	Production	Hours	Production	Hours
Production	0.9555	–0.0109	0.9823	0.0225
Hours	–0.0179	0.6975	0.4634	0.6028

The roots associated with the coefficient matrix for Italy are obviously about 0.956 and 0.698. For the U.S. matrix they are 1.008 and 0.577. Clearly a single-unit-root hypothesis is *also* consistent with the data, indeed even more consistent with it (in terms of likelihood) than the two-unit-roots hypothesis. Hours in the United States are estimated as nonstationary only because of cointegration with productivity. In Italy they are estimated as stationary, with the paper's identifying assumption very nearly satisfied by the reduced form, taking the nontechnology shock as simply the productivity innovation.

It is possible to deal directly with the uncertainty about whether roots are exactly 1 and how many roots are nonstationary. Bayesian inference in these models has no need for preliminary tests to determine stationarity or for identifying cointegrating vectors and conditioning inference on them. From a Bayesian perspective, all scientific reporting of inference is best regarded as helping readers to understand the shape of the likelihood function, on which any decision making use of the results ought to be based. The likelihood function for these models shows no special behavior as we cross from stationary to nonstationary regions of the parameter space.

I have reestimated the paper's structural VAR using priors that pull toward unit root behavior with varying intensity and toward forms with one or two unit roots. In most countries, the qualitative results are similar to those shown in the paper, with unambiguously negative responses of hours to identified technology shocks in the same countries where the paper finds them. In the United States, though, results are sensitive to how insistent one is about there being two distinct unit roots.

The prior I use (documented in Sims and Zha, 1998, and also in the comments to the code in mgnldnsty.R or mgnldnsty.m available at sims.princeton.edu/yftp/VARtools) has one component indexed by the parameter λ that pulls estimates toward at least one unit root and zero constant term, or else toward stationarity (with a nontrivial constant). Another component, indexed by μ, pulls toward independent unit roots in all variables. When μ is even moderately large, the posterior peaks at two roots very close to 1. But for any given value of μ, as λ increases, the estimates eventually flip to showing a positive response of hours to productivity shocks at all horizons. Figure 1 shows the posterior modal response for $\lambda = 0.5$, $\mu = 0$, together with the 90 percent error band. The error band is not so different from that shown in Galí's paper, but the location of the modal response is very different. This prior is very weak, but not so weak as to imply low posterior odds. The marginal data density corresponding to the plot is within a factor of 10 of the highest marginal density I have found by varying parameters of the prior. It is possible to get the same pattern in the modal responses, higher marginal data densities, and narrower implied error bands, by tightening up the prior somewhat.

Figure 1

Response of Employment to Technology Shock: λ = 0.5, μ = 0

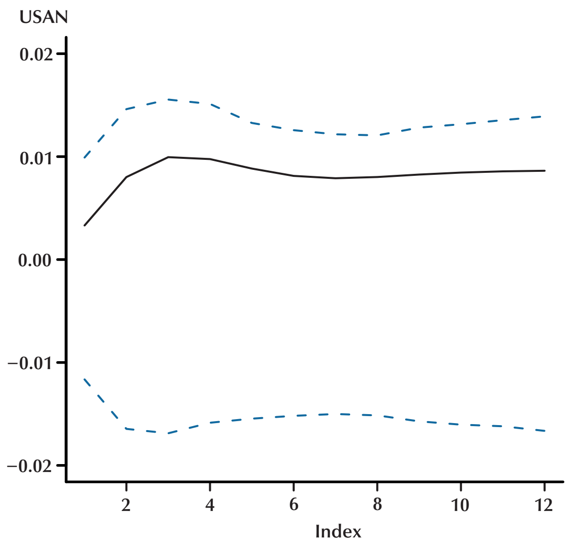
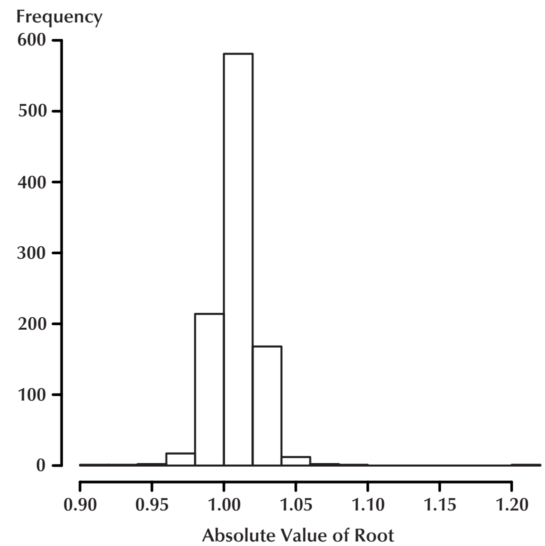


Figure 2

Largest Root, λ = 0.5, μ = 0



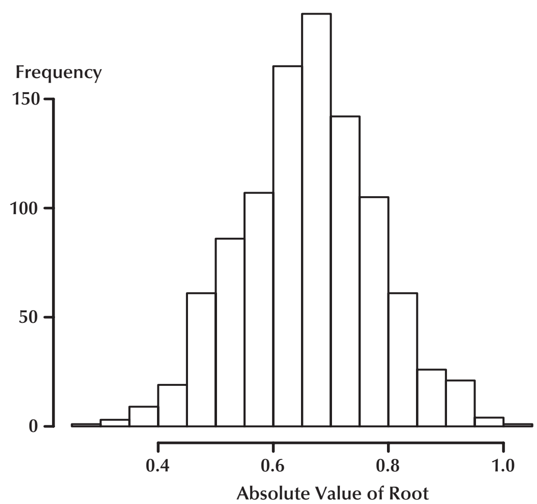
The degree to which the data leave the sizes of roots indeterminate can be seen from the posterior probability density functions for the absolute values of the smallest and largest roots, shown in Figures 2 and 3, computed with the same fairly diffuse prior as was used to generate Figure 1.

These results only strengthen a conclusion already available from the impulse response graphs at the end of Galí's paper. Those figures show that the response of hours to a technology shock is estimated as significantly positive in one country (Japan), significantly negative in two countries (United Kingdom and Italy), and indeterminate in four (United States, Canada, France, and Germany). Reworking the U.S. data allowing for uncertainty in the number of roots has made the U.S. results even more uncertain and, with one reasonable prior, made the modal response positive rather than negative.

In fact, the responses of gross domestic product (GDP) to a technology shock are equally unstable across countries. Four countries (Italy, United States, Canada, and Germany) show indeterminate responses of GDP to a technology shock, one shows a negative response (United Kingdom), and two show a positive response (France and Japan).

Figure 3

Second-Largest Root, λ = 0.5, μ = 0



It is natural then, I would say, to question whether the paper's methodology is isolating the same two structural shock in all of these countries. Note that at least for the United States and Europe, Smets and Wouters found in their larger model patterns of impulse responses that were quite stable across countries.

CONCLUSION

The paper is carefully done and thought-provoking. To get the most out of it, one should not let the clash between one-dimensional RBC models and two-dimensional SVAR models that occupies the foreground of the paper hide the background issues that the paper illuminates:

- Uncertainty about stationarity matters.
- If we are to integrate our modeling of long- and short-run macro-dynamics, it appears we need to go beyond one- and two-dimensional models.

REFERENCES

Chari, V.V.; Kehoe, Patrick J. and McGrattan, Ellen. "An Economic Test of Structural VARs." Federal Reserve Bank of Minneapolis, Research Department Staff Report 345, 2004.

Christiano, Lawrence; Eichenbaum, Martin and Vigfusson, Robert. "What Happens after a Technology Shock?" NBER Working Paper No. 9819, National Bureau of Economic Research, 2003.

Galí, Jordi. "Technology, Employment, and the Business Cycle: Do Technology Shocks Explain Aggregate Fluctuations?" *American Economic Review*, 1999, *89*(1), pp. 249-71.

Galí, Jordi. "Trends in Hours, Balanced Growth, and the Role of Technology in the Business Cycle." Federal Reserve Bank of St. Louis *Review*, July/August 2005, *87*(4), pp. 459-86.

Galí, Jordi and Rabanal, Pau. "Technology Shocks and Aggregate Fluctuations: How Well Does the Real Business Cycle Model Fit Postwar U.S. Data?" NBER Working Paper No. 10636, National Bureau of Economic Research, 2004 (*NBER Macroeconomics Annual*, forthcoming).

Shapiro, Matthew D. and Watson, Mark W. "Sources of Business Cycle Fluctuations." *NBER Macroeconomics Annual 1988*. Cambridge, MA: MIT Press, 1989, pp. 111-48.

Sims, Christopher A. and Zha, Tao. "Bayesian Methods for Dynamic Multivariate Models." *International Economic Review*, November 1998, *39*(4), pp. 949-68.

Smets, Frank and Wouters, Raf. "An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area." *Journal of the European Economic Association*, September 2003a, *1*, pp. 1123-75.

Smets, Frank and Wouters, Raf. "Shocks and Frictions in U.S. Business Cycles: A Bayesian DSGE Approach." Discussion paper, European Central Bank and National Bank of Belgium, May 2003b.

The Cyclicality of Hires, Separations, and Job-to-Job Transitions

Robert Shimer

This paper measures the job-finding, separation, and job-to-job transition rates in the United States from 1948 to 2004. The job-finding and job-to-job transition rates are strongly procyclical and the separation rate is nearly acyclical, especially since 1985. The author develops a simple model in which unemployed workers search for jobs and employed workers search for better jobs. The model predicts that an increase in either the job-finding rate or the separation rate raises the job-to-job transition rate, which is qualitatively and quantitatively consistent with the available evidence. In contrast, if the job-finding rate were acyclical and the separation rate countercyclical, as is the conventional wisdom, the model predicts that the job-to-job transition rate would be counterfactually countercyclical.

Federal Reserve Bank of St. Louis *Review*, July/August 2005, *87*(4), pp. 493-507.

I measure the job-finding and separation rates in the United States from 1948 to 2004 and find that there are substantial fluctuations in the job-finding probability—the monthly probability that a typical unemployed worker finds a job—at business cycle frequencies, whereas the separation probability—the monthly probability that a typical employed worker becomes unemployed—is comparatively acyclical (Figure 1). This finding is particularly true in the past two decades, a period in which the separation probability has steadily declined despite two spikes in the unemployment rate.

I then put these measures of the job-finding and separation probabilities into a simple model of job-to-job transitions. I assume employed workers continuously search for better employment opportunities: They experience a measured job-to-job transition either when they find a better job or when they are forced to leave their previous job but manage to find a new one before they are counted as unemployed. I show that an increase in the job-finding rate or an increase in the separation rate raises the job-to-job transition rate. Therefore, when I feed the measured time series for the job-finding and separation rates into this simple model, I predict that the job-to-job transition rate should be procyclical (Figure 4). This is quantitatively consistent with two direct measures of the job-to-job transition rate (Figures 5 and 6). In contrast, if separations were countercyclical and the job-finding rate acyclical, the basic model would predict a countercyclical job-to-job transition rate.

My findings that the job-finding rate is strongly procyclical and the separation rate is nearly acyclical oppose the conventional wisdom that recessions are primarily characterized by a high separation rate. In their 1996 book, Davis, Haltiwanger, and Schuh, building on evidence developed by Davis and Haltiwanger (1990 and 1992), conclude that evidence from the U.S. manu-

Robert Shimer is a professor of economics at the University of Chicago and a research associate at the National Bureau of Economic Research. The author is grateful for comments from Fernando Alvarez, Gadi Barlevy, Francesco Belviso, Robert Hall, and Randall Wright and from seminar participants at the Bank of Italy, Bocconi University, the Federal Reserve Bank of Chicago, Harvard University, the Federal Reserve Bank of St. Louis, and the University of Texas–Austin on an earlier version of this paper. Some of the ideas in that version are now contained in Shimer (2005b). The current paper was supported by grants from the National Science Foundation and the Sloan Foundation.

facturing sector indicates that "job destruction rises dramatically during recessions, whereas job creation initially declines by a relatively modest amount" (p. 34). Blanchard and Diamond (1990, p. 87) reach a similar conclusion from their analysis of both worker and job flows: "The amplitude of fluctuations in the flow out of employment is larger than that of the flow into employment. This, in turn, implies a much larger amplitude of the underlying fluctuations in job destruction than of job creation."[1]

The development of macroeconomic models of the labor market has been profoundly affected by the conventional wisdom,[2] but a series of recent papers by Hall (2004; 2005a,b) and Shimer (2005a,b) examine a variety of new data sets and question the prevailing view. For example, Hall (2004) writes that, "in the modern U.S. economy, recessions are not times of unusual job loss. New data on separations show them to be remarkably constant from peak to trough. Bursts of job loss had some role in earlier recessions, but are still mostly a side issue for the reason just mentioned—a burst is quickly reabsorbed because of high job-finding rates."

Relative to this new empirical literature, this paper's main contribution is its focus on job-to-job transitions. Burdett (1978) developed the first model of search by both unemployed and employed workers, showing that an unemployed worker uses a reservation strategy, accepting any job whose quality exceeds a lower bound, while an employed worker takes any job that is better than his current one. Pissarides (1994) and Burdett and Mortensen (1998) placed on-the-job search models into an equilibrium framework. I am aware of three recent analyses of on-the-job search in the presence of economic fluctuations: Barlevy (2002), who emphasizes that a decrease in the job-to-job transition rate during recessions leaves workers in less-productive jobs; Kraus and Lubik (2004), who show that job-to-job transitions can give rise to vacancy chains (Akerlof, Rose, and Yellen, 1988), amplifying fluctuations in the vacancy-unemployment ratio; and Nagypál (2004a), who examines a firm's choice of whether to hire an employed or unemployed worker and shows that a preference for hiring employed workers can also generate large fluctuations in the vacancy-unemployment ratio.

The methodology in this paper closely follows Shimer (2005b). Both papers emphasize that time aggregation leads to an overstatement of the cyclicality of the separation rate and offer a correction. Shimer (2005b) compares the proposed measure of the job-finding and separation rates with alternatives in the literature: It argues that cyclical changes in the composition of the unemployed population do not drive fluctuations in the job-finding and separation rates; and it shows that cyclical movements of workers in and out of the labor force are also unimportant. The present paper focuses on developing a simple model of job-to-job transitions. It provides several different measures of the job-to-job transition rate and shows that, if in fact the job-finding rate is procyclical and the separation rate is acyclical, the model can account quantitatively for the behavior of the job-to-job transition rate.

When possible, I use readily available aggregate data so comparable measures can easily be constructed for particular industries or groups of workers or for other countries.[3] I show how data on unemployment duration constructed by the Bureau of Labor Statistics (BLS) from the Current Population Survey (CPS) can be used to measure both the job-finding and separation rates. On the other hand, measures of the job-to-job transition rate necessitate using the underlying microeconomic data.

The next section provides measures of the job-finding and separation rates, accounting carefully for time aggregation. The third section develops a simple model of job-to-job transitions and discusses the theoretical response to an innovation in the job-finding or separation rate. It then develops several measures of the job-to-job tran-

[1] For some skeptical views on these stylized facts, see Boeri (1996) and Foote (1998).

[2] An incomplete list of well-known papers that embrace the conventional wisdom includes Mortensen and Pissarides (1994), Caballero and Hammour (1994), Hall (1995), Ramey and Watson (1997), Cole and Rogerson (1999), den Haan, Ramey, and Watson, (2000), and Pries (2004).

[3] But I have constructed these measures only for the United States.

sition rate and argues that all are strongly procyclical and quantitatively consistent with the predictions of the simple model.

MEASURING THE JOB-FINDING AND SEPARATION RATES

This section develops theoretical measures of the job-finding rate for unemployed workers, f_t, and the separation rate for employed workers, s_t, during period t and uses publicly available data from the CPS to measure the two transition rates. The job-finding rate is strongly procyclical, whereas the separation rate is less cyclical and explains little of the overall fluctuations in employment and unemployment, particularly during the past two decades.

Theory

I make two critical assumptions in this section; see Shimer (2005b) for evidence that they are good approximations of reality. First, I ignore movements in and out of the labor force, so workers simply move back and forth between employment and unemployment, where the latter is defined as the state of active job search. Second, I assume that all unemployed workers find a job at rate f_t and all employed workers lose a job at rate s_t during period t. In particular, I ignore any heterogeneity or duration dependence that makes some unemployed workers more likely to find and some employed workers less likely to lose a job within the period.

I model a continuous time environment in which data are available only at discrete dates $t \in \{0,1,2,\ldots\}$. I refer to the interval $[t, t+1)$ as "period t"; fix $t \in \{0,1,2,\ldots\}$ and let $\tau \in [0,1]$ denote the time elapsed since the previous measurement date; and let $e_{t+\tau}$ denote the number of employed workers at time $t+\tau$, $u_{t+\tau}$ the number of unemployed workers at time $t+\tau$, and $u_t^s(\tau)$ as "short-term unemployment," workers who are unemployed at time $t+\tau$ but were employed at some time $t' \in [t, t+\tau]$. Note that $u_t^s(0) = 0$ for all t. It is convenient to define $u_{t+1}^s \equiv u_t^s(1)$ as the total amount of short-term unemployment at the end of period t. Because I cannot observe within-period variation in the job-finding or separation rate using available data, I assume that these are constant within periods. Let $f_t \geq 0$ and $s_t \geq 0$ denote the rate at which an unemployed worker finds a job and an employed worker loses a job, respectively, during period $t \in \{0,1,2,\ldots\}$. Throughout this paper, I refer to f_t and s_t as the job-finding and separation *rates*, respectively, and to $F_t \equiv 1 - e^{-f_t} \in [0,1]$ and $S_t \equiv 1 - e^{-s_t} \in [0,1]$ as the corresponding *probabilities*, respectively (i.e., F_t is the probability that a worker who begins period t unemployed finds at least one job during the period and similarly for S_t).

For $t \in \{0,1,2,\ldots\}$ and $\tau \in [0,1)$, unemployment and short-term unemployment evolve according to

$$\dot{u}_{t+\tau} = e_{t+\tau}s_t - u_{t+\tau}f_t \tag{1}$$

and

$$\dot{u}_t^s(\tau) = e_{t+\tau}s_t - u_t^s(\tau)f_t. \tag{2}$$

Unemployment increases when employed workers separate from their jobs and decreases when unemployed workers find jobs, whereas short-term unemployment increases when employed workers separate from their jobs and decreases when short-term unemployed workers find jobs. Eliminate $e_{t+\tau}s_t$ between these equations to get

$$\dot{u}_{t+\tau} = \dot{u}_t^s(\tau) - \left(u_{t+\tau} - u_t^s(\tau)\right)f_t,$$

for $\tau \in [0,1)$. By construction, $u_t^s(0) = 0$, so given an initial condition for u_t, this can be solved for u_{t+1} and $u_{t+1}^s \equiv u_t^s(1)$:

$$u_{t+1} = \left(1 - F_t\right)u_t + u_{t+1}^s. \tag{3}$$

The number of unemployed workers at date $t+1$ is equal to the number of unemployed workers at date t who do not find a job during the period (fraction $1 - F_t = e^{-f_t}$) plus the u_{t+1}^s short-term unemployed workers, those who are unemployed at date $t+1$ but held a job at some point during period t. Given measures of unemployment and short-term unemployment, it is straightforward to back-out the job-finding probability from equation (3)

By assumption, the labor force, $l_t = e_t + u_t$, is constant during period t. This enables me to solve the differential equations (1) forward to obtain an implicit expression for the separation rate:

$$u_{t+1} = \frac{(1-e^{-f_t-s_t})l_t s_t}{f_t+s_t} + e^{-f_t-s_t}u_t. \tag{4}$$

Because $l_t > u_t$, the right-hand side of this expression is increasing in s_t. Given the job-finding rate from equation (3) and data on unemployment and employment, equation (4) uniquely defines the separation rate, s_t.[4]

To understand equation (4), note first that if unemployment is constant during period t, this reduces to a familiar formula, $u_t/l_t = s_t/(s_t + f_t)$, so the unemployment rate is determined by the ratio of the separation rate to the job-finding rate. Out of steady state, it helps to compare equation (4) with a discrete time model where there is no possibility of both finding and losing a job within a period. In this case,

$$u_{t+1} = S_t e_t + \left(1-F_t\right)u_t. \tag{5}$$

A fraction S_t of employed workers lose their job and a fraction F_t of unemployed workers find a job during period t, determining the unemployment rate at the start of period $t+1$. When the time period is sufficiently short, equation (4) converges to this simple expression. But with longer time periods, equation (4) allows workers to lose a job and find a new one, or vice versa, within the period. The distinction is quantitatively important for measuring both the level and cyclicality of the separation rate. When equation (3) indicates that the job-finding rate is high, a worker who loses her job is more likely to find a new one without experiencing a measured spell of unemployment. These separations are counted in equation (4) but missed in equation (5), so the latter formula yields an artificially negative correlation between the job-finding and separation rates—that is, a time aggregation bias.

[4] A previous version of this paper proposed measuring the separation rate as $u^s_{t+1}f_t/e_t(1-e^{-f_t})$, because a fraction $(1-e^{-f_t})/f_t$ of the workers who lose their job during the period are still unemployed at the next measurement date. This is virtually identical to the measure of s_t that I now use.

Measurement

Since 1948, the BLS has published monthly data on employment, unemployment, and unemployment duration based on the CPS. The measures of the number of employed and unemployed workers are standard, and I use these to quantify e_t and u_t. The survey also asks unemployed workers how long they have been unemployed, and the BLS tabulates the number of unemployed workers with zero to four weeks' duration. With an adjustment for the CPS redesign in January 1994, discussed in the appendix, I use this as my measure of short-term unemployment, u^s_t.[5]

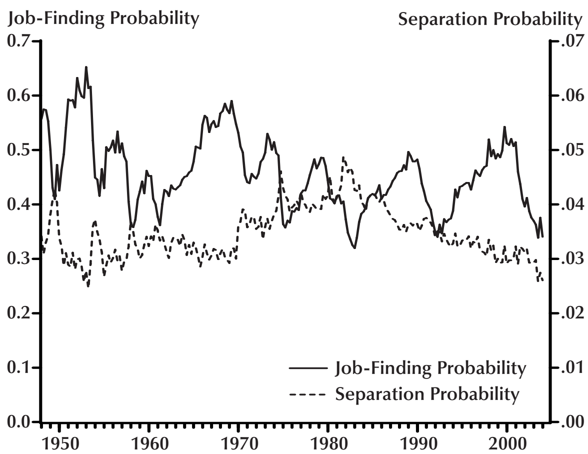
Figure 1 shows the time series for the job-finding probability, $F_t = 1 - e^{-f_t}$, and separation probability, $S_t = 1 - e^{-s_t}$, constructed according to equations (3) and (4) from 1948 to 2004. Several facts stand out. First, the job-finding probability is high, averaging 46 percentage points over the postwar period. Second, the job-finding probability is variable, falling by about 40 log points from peak to trough during recent decades. Third, the separation probability averaged 3.5 percentage points during the same period and was somewhat less volatile, particularly in recent years.

To examine the cyclicality of the job-finding and separation rates, recall that if unemployment is constant, $u_t = u_{t+1}$, equation (4) implies that the unemployment rate is $u_t/l_t = s_t/(s_t + f_t)$. In fact, this is a very good approximation. In monthly data, the correlation between the period t "steady-state unemployment rate," $s_t/(s_t + f_t)$, and the end-of-period actual unemployment rate, u_{t+1}/l_{t+1}, is almost 0.99; so, the current job-finding and separation rates determine future unemployment rates. I use this strong relationship to distinguish between the importance of fluctuations in the job-finding and separation rates for determining fluctuations in unemployment. Let $\bar{f}$ and $\bar{s}$ denote the average values of f_t and s_t during the sample period and compute $\bar{s}/(\bar{s} + f_t)$ and $s_t/(s_t + \bar{f})$ as measures of the contributions of fluctuations in the job-finding and separation rates to overall fluctuations in the unemployment rate.

[5] These data can be downloaded from the BLS web site (www.bls.gov) or from the Federal Reserve Economic Database (FRED®II) of the Federal Reserve Bank of St. Louis: http://research.stlouisfed.org/fred2/.

Figure 1

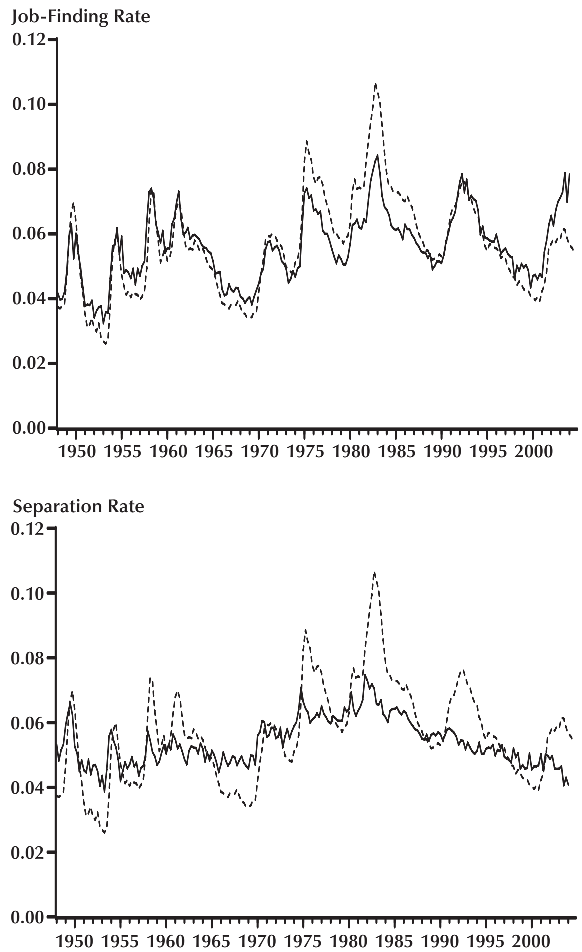
Job-Finding and Separation Probabilities



NOTE: Data, quarterly averages of monthly data, are from 1948:Q1–2004:Q1. The job-finding probability $F_t=1-e^{-f_t}$ is constructed from unemployment and short-term unemployment according to equation (3). The separation probability $S_t=1-e^{-s_t}$ is constructed from employment, unemployment, and the job-finding rate according to equation (4). Employment, unemployment, and short-term unemployment data are constructed by the BLS from the CPS and are seasonally adjusted.

Figure 2

Contribution of Fluctuations in the Job-Finding and Separation Rates to Fluctuations in the Unemployment Rate



NOTE: Data, quarterly averages of monthly data, are from 1948:Q1–2004:Q1. The job-finding rate, f_t, is constructed from unemployment and short-term unemployment according to equation (3). The separation rate s_t is constructed from employment, unemployment, and the job-finding rate according to equation (4). The top panel shows $\bar{s}/(\bar{s}+f_t)$ and the bottom panel shows $s_t/(s_t+\bar{f})$ as a solid line, with the unemployment rate graphed as a dashed line for comparison in each panel. Employment, unemployment, and short-term unemployment data are constructed by the BLS from the CPS and are seasonally adjusted.

The top panel in Figure 2 shows that a decline in the job-finding rate, f_t, contributed to every increase in the unemployment rate during the postwar period. The bottom panel shows that, from 1948 to 1985, the separation rate tended to move with the unemployment rate, although it rarely explained more than half the fluctuations in unemployment. In the past two decades, however, the separation rate has varied little over the business cycle. One way to quantify this is to look at the comovement of the detrended data.[6] Over the entire postwar period, the correlation between the cyclical components of u_{t+1}/l_{t+1} and $\bar{s}/(\bar{s}+f_t)$ is 0.96, whereas the correlation between u_{t+1}/l_{t+1} and $s_t/(s_t+\bar{f})$ is 0.71. The latter correlation has

[6] I time-aggregate the underlying monthly data to get quarterly averages, removing substantial low-frequency fluctuations that likely reflect measurement error in the CPS. I then detrend the quarterly data using an HP filter with smoothing parameter 10^5. This is a much lower-frequency filter than is commonly used in business cycle analyses of quarterly data. A standard filter seems to remove much of the cyclical volatility in the variable of interest.

fallen to 0.20 since 1986. Moreover, $\bar{s}/(\bar{s}+f_t)$ is relatively volatile, with a cyclical standard deviation equal to 0.99 times that of u_{t+1}/l_{t+1}. The relative standard deviation of $s_t/(s_t+\bar{f})$ is just 0.28.

JOB-TO-JOB TRANSITIONS

I now extend the basic model of transitions between employment and unemployment to allow employed workers to search for better jobs. I show analytically that a permanent increase in the job-finding rate or in the separation rate raises the job-to-job transition rate. I also show numerically that a transitory increase has a similar effect. This implies that if the results described in the previous section are correct—that the job-finding rate is strongly procyclical and the separation rate is only weakly countercyclical—the job-to-job transition rate should be procyclical in the United States. I then confirm this prediction, both qualitatively and quantitatively, using a variety of data sources.

Theory

As before, assume that unemployed workers find a job at rate f_t during month t and employed workers lose their job at rate s_t. In addition, employed workers find an alternative job at rate f_t^e. To explain why workers switch jobs, I assume jobs are of different "quality," y, an index summarizing all of the job's pecuniary and nonpecuniary aspects. An unemployed worker accepts any job,[7] whereas an employed worker accepts a new job only if its quality y' exceeds the previous job's quality, y. The critical assumption is that when a worker finds a job, the quality is drawn from a time-invariant continuous distribution, $Z(y)$, with support $[0,\bar{y}]$. This gives a simple model of job-to-job transitions in which workers switch jobs whenever they have an opportunity to improve their job quality.

To compute the theoretical job-to-job transition rate, it is necessary to keep track of the distribution of employed workers across job qualities. Let $\tilde{G}_{t+\tau}(y)$ denote the fraction of employed workers whose job quality is less than y at date $t+\tau$. Then $\tilde{G}_{t+\tau}(y)\, e_{t+\tau}$ is the total number of employed workers in jobs worse than y. For any $t+\tau$, this evolves according to

$$\frac{d\tilde{G}_{t+\tau}(y)e_{t+\tau}}{d\tau}$$
$$=u_{t+\tau}f_{t+\tau}Z(y)-\tilde{G}_{t+\tau}(y)e_{t+\tau}\left(s_{t+\tau}+f_{t+\tau}^{e}\left(1-Z(y)\right)\right).$$

It increases when the $u_{t+\tau}$ unemployed workers find a job, at rate $f_{t+\tau}$, and the job has quality below y. It decreases when one of the $\tilde{G}_{t+\tau}(y)\, e_{t+\tau}$ workers employed in a job with quality below y either suffers a separation (rate $s_{t+\tau}$) or finds a new job with quality above y (rate $f_{t+\tau}^e(1-Z(y))$). Because employment increases when unemployed workers find jobs and decreases when employed workers lose jobs, $\dot{e}_{t+\tau}=u_{t+\tau}f_{t+\tau}-e_{t+\tau}s_{t+\tau}$, the previous equation may be rewritten as

(6)
$$\dot{\tilde{G}}_{t+\tau}(y)=$$
$$\frac{u_{t+\tau}f_{t+\tau}}{e_{t+\tau}}\left(Z(y)-\tilde{G}_{t+\tau}(y)\right)-\tilde{G}_{t+\tau}(y)f_{t+\tau}^{e}\left(1-Z(y)\right).$$

I will say that a voluntary job-to-job transition occurs when a worker in a job with quality y finds a better job, giving an instantaneous flow,

(7)
$$\begin{aligned} j_{t+\tau}^{v} &= f_{t+\tau}^{e}e_{t+\tau}\int_0^{\bar{y}}\left(1-Z(y)\right)\tilde{G}_{t+\tau}'(y)dy \\ &= f_{t+\tau}^{e}e_{t+\tau}\int_0^{\bar{y}}\tilde{G}_{t+\tau}(y)Z'(y)dy, \end{aligned}$$

where the second equality uses integration by parts. The total number of voluntary job changers during period t is

(8)
$$J_t^v=\int_0^1 j_{t+\tau}^v d\tau .$$

The main difficulty in measuring J_t^v using this equation is that the quality distribution, Z, is unobservable. Fortunately, this is unimportant. Rather than indexing a job opportunity by its quality, y, drawn from the distribution, Z, I represent it by its percentile in the quality distribution, z, which by definition is distributed uniformly on [0,1]. In addition, I define $G_t(Z(y))\equiv\tilde{G}_t(y)$ for all y. Then for $t\in\{0,1,2,\ldots\}$ and $\tau\in[0,1)$, the distribution of workers' normalized quality

[7] Burdett (1978) shows that an unemployed worker's acceptance threshold will generally depend on labor market conditions. To keep the analysis simple, I ignore that possibility here.

and the voluntary job-to-job transition rate satisfy

$$(9)\quad \dot{G}_{t+\tau}(z)=\frac{u_{t+\tau}f_t}{e_{t+\tau}}\left(z-G_{t+\tau}(z)\right)-G_{t+\tau}(z)f_t^e(1-z)$$

$$(10)\qquad J_t^v=f_t^e\int_0^1 e_{t+\tau}\left(\int_0^1 G_{t+\tau}(z)dz\right)d\tau,$$

where I use the standard assumption that the job-finding rates, $f_{t+\tau}$ and $f_{t+\tau}^e$, are constant during period t to simplify these expressions.

Given an initial guess at the distribution of G and data on unemployment, u_t, employment, e_t, and job-finding rates, f_t and f_t^e, it is possible to compute future distributions using equation (9). More precisely, I start each period t with initial values for u_t and e_t (measured by the BLS from the CPS) and $G_t(z)$ (computed in the previous period). Second, I solve (1) to compute $u_{t+\tau}$ and $e_{t+\tau}\equiv u_t+e_t-u_{t+\tau}$, $\tau\in[0,1)$. Third, I use equation (9) to compute $G_{t+\tau}(z)$, $\tau\in[0,1)$, on a grid of points $z\in\{0, 0.01, 0.02, \ldots, 1\}$. This computation has two purposes: $G_{t+1}(z)$ serves as initial condition in the next period, and I can approximate the entire function $G_{t+\tau}(z)$ with a cubic spline, calculating $\int_0^1 G_{t+\tau}(z)dz$ at times $\tau\in\{0, 0.1, 0.2, \ldots, 1\}$. Again, using a cubic spline approximation, now along the time dimension, I can finally evaluate equation (10) to yield J_t^v.

Time aggregation introduces a second type of job-to-job transition, one that occurs when a worker separates from her job but manages to find a new one before the next survey date. At time $t+\tau$, there are $u_t^s(\tau)$ workers who have separated from their jobs since the previous survey date, each of whom finds a job at rate f_t. This means the measured "involuntary" job-to-job transition rate between months t and $t+1$ is

$$(11)\qquad J_t^i=f_t\int_0^1 u_t^s(\tau)d\tau.$$

This rate is easy to measure using the variables constructed in the second section, "Measuring the Job-Finding and Separation Rates."

Finally, the total number of workers who switch employers during the preceding month is $J_t=J_t^v+J_t^i$:

$$(12)\quad J_t=f_t^e\int_0^1 e_{t+\tau}\left(\int_0^1 G_{t+\tau}(z)dz\right)d\tau+f_t\int_0^1 u_t^s(\tau)d\tau.$$

The job-to-job transition rate is the ratio of this to employment, J_t/e_t.

Comparative Statics and Impulse Response

In steady state, it is possible to solve equation (9) for $G(z)$, simplifying with the steady-state unemployment condition $uf=es$:

$$G(z)=\frac{sz}{s+f^e(1-z)}.$$

I substitute this into (10) to obtain the voluntary job-to-job transition rate:

$$\frac{J^v}{e}=s\left(\frac{f^e+s}{f^e}\log\left(\frac{f^e+s}{f^e}\right)-1\right).$$

This rate is independent of the job-finding rate for unemployed workers, f, but is increasing in both the job-finding rate for employed workers, f^e, and in the separation rate, s. That the voluntary job-to-job transition rate is increasing in the job-finding rate for employed workers is intuitive. To understand the comparative static with respect to the separation rate, consider an extreme economy without separations. Eventually all workers will find the best possible job, $z=1$, and there will be no more job-to-job transitions. A higher separation rate pushes more workers down the job ladder, which raises the possibility of voluntary job-to-job transitions.

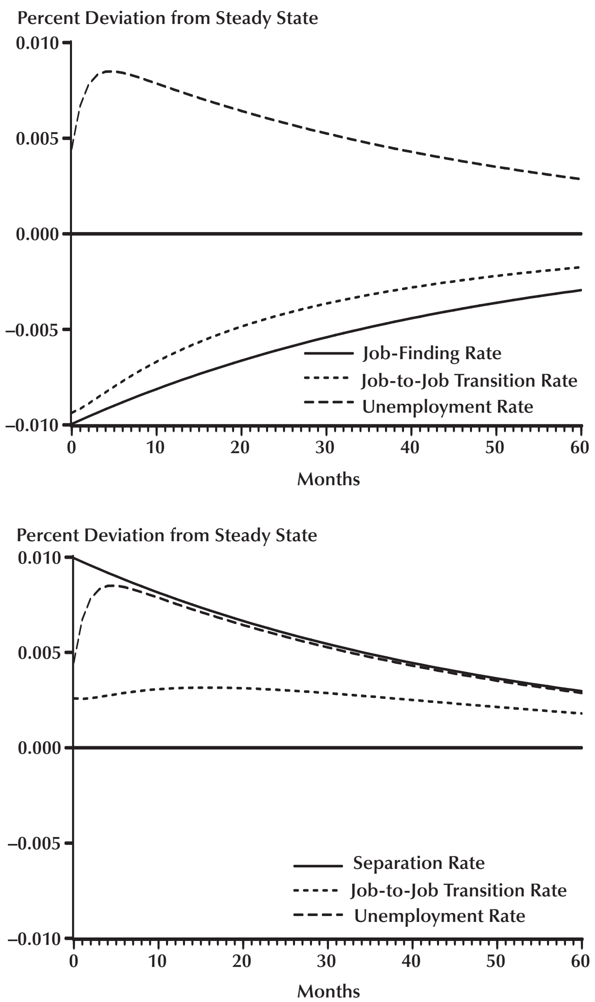
Still in steady state, equation (2) implies $\dot{u}_t^s(\tau)=e_ts_t-u_t^s(\tau)f_t$. I use the initial condition $u_t^s(0)=0$ and integrate over $\tau\in[0,1]$ to get an analytic expression for $\int_0^1 u_t^s(\tau)d\tau$, which can be substituted into equation (11). This calculation implies that the steady-state involuntary job-to-job transition rate is

$$\frac{J^i}{e}=s\left(1-\frac{1-e^{-f}}{f}\right).$$

This rate is unaffected by the job-finding rate for employed workers but is increasing in both the separation rate and the job-finding rate for unemployed workers. A higher separation rate places more workers in a position to undertake an involuntary job-to-job transition, whereas a higher job-finding rate permits more of these

Figure 3

Theoretical Response of Unemployment and the Job-to-Job Transition Rate to a Shock to the Job-Finding or Separation Rate



workers to do so. In summary, the steady-state job-to-job transition rate,

$$\frac{J}{e} = s\left(\frac{f^e + s}{f^e}\log\left(\frac{f^e + s}{f^e}\right) - \frac{1 - e^{-f}}{f}\right),$$

is increasing in the separation rate and both job-finding rates.

The behavior of the economy out of steady state is similar to that suggested by the comparative statics. An increase in the separation rate or a reduction in the job-finding rate both raise the unemployment rate, but their effects on the job-to-job transition rate are distinct. An increase in the separation rate initially raises the involuntary job-to-job transition rate as some displaced workers find their way back into a new job within the period. Additionally, higher separation rates in the recent past raise the voluntary job-to-job transition rate as previously displaced workers go through multiple jobs before finding a very good match. Thus, the job-to-job transition rate tends to increase and remain high following a transitory increase in the separation rate. In contrast, a decrease in the job-finding rate reduces both voluntary and involuntary job-to-job transitions because fewer employed or recently displaced workers find a worthwhile job opportunity within the period.

To quantify these effects, I consider an economy that is initially in steady state with a job-finding rate of $f_{-1} = 0.61$ and a separation rate of $s_{-1} = 0.035$, equal to the average U.S. monthly transition rates from 1948 to 2004. I also set $f^e_{-1} = 0.15 f_{-1}$, a reasonable value for the relative efficiency of on-the-job search, as I will discuss further. Assume that the parameters of the economy have been constant for a while, so the unemployment rate is at its steady-state value, $u_{-1} = 0.054$, and the distribution of employed workers across job quality is also in steady state. Consider two experiments in this model economy: a 1 percent increase in the separation rate that reverts to steady state at 2 percent per month, so $s_t = 0.03535 \times 0.98^t + 0.35 \times (1 - 0.98^t)$ for $t \in \{0,1, 2,\ldots\}$, with $f_t = 0.61$ for all t; or a 1 percent, mean-reverting decrease in the job-finding rate, so $f_t = 0.60396 \times 0.98^t + 0.61 \times (1 - 0.98^t)$ for $t \in \{0,1, 2,\ldots\}$, with $s_t = 0.035$ for all t. In both cases, I let $f^e_t = 0.15 f_t$ for all t.

Figure 3 shows the behavior of unemployment and the job-to-job transition rate in response to these shocks. By construction, unemployment behaves identically, but the cyclicality of the job-to-job transition rate depends on the nature of the shock. The top panel shows that the job-to-job transition rate closely tracks a transitory decrease in the job-finding rate; the bottom panel shows that the job-to-job transition rate fluctuates less and is more persistent in response to a transitory increase in the separation rate. Most important, if cyclical volatility in unemployment is primarily due to fluctuations in the job-finding rate, the model predicts that the job-to-job transition rate should be procyclical (i.e., negatively correlated with unemployment). If it is primarily due to fluctuations in the separation rate, the job-to-job transition rate should be countercyclical (positively correlated with unemployment).

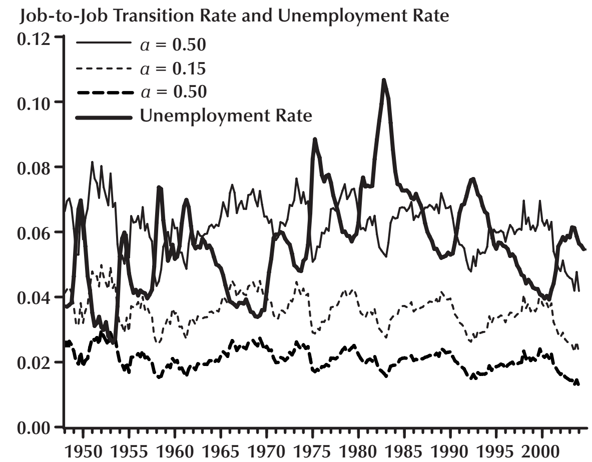
Measurement

To test this theory, I compare the job-to-job transition rate predicted by this simple model with direct evidence on the empirical transition rate. Because the job-finding rate is strongly procyclical and the separation rate is weakly countercyclical, it is not obvious a priori whether the theory predicts a procyclical or countercyclical job-to-job transition rate. To answer this, I feed into the model monthly employment and unemployment measures from the CPS (e_t and u_t, respectively); the monthly job-finding rate, f_t, computed from equation (3); and the monthly separation rate, s_t, computed from equation (4). I also fix $f_t^e = af_t$ and examine three values for the relative efficiency of on-the-job search, $a \in \{0.50, 0.15, 0.05\}$. Figure 4 shows the resulting series for the job-to-job transition rate, constructed according to equation (12), with the unemployment rate plotted for comparison. The theoretical job-to-job transition rate is strongly negatively correlated with unemployment, regardless of the value of a. After removing a low-frequency trend, the correlation between the job-to-job transition rate and the unemployment rate ranges from –0.89 with $a = 0.05$ to –0.80 with $a = 0.50$, and the relative standard deviation lies between 0.65 with $a = 0.05$ and 0.53 with $a = 0.50$. For example, the theory suggests that the job-to-job transition rate should have fallen by between 0.31 and 0.44 log points during 2001 and 2002, a period during which the job-finding rate fell and the separation rate remained roughly constant.

There is no ideal empirical measure of the job-to-job transition rate to test this prediction, so I rely on three imperfect measures. The first methodology, pioneered by Fallick and Fleischman (2004) and recently employed by Nagypál (2004b) examines a question in the public use microeconomic data from the monthly CPS.[8] Since the switch to dependent interviewing in 1994, the survey has asked the following question of respondents who continue to be employed in consecutive months: "Last month, it was reported that you worked for *x*. Do you still work for *x* (at your main job)?" I use the fraction of employed workers who answer this question negatively, weighted by the CPS final weights, to compute the empirical job-to-job transition rate. A potential shortcoming of this method is that no individual is permitted to experience multiple job-to-job movements within a month, a possibility that is nonnegligible when the job-finding rate is high. More importantly, because the relevant question has been asked only since 1994, it is possible to

Figure 4

Theoretical Job-to-Job Transition Rate, Constructed According to Equation (12)

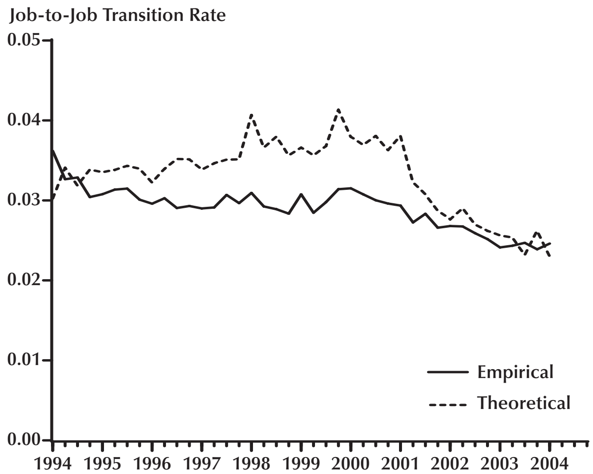


NOTE: Data, quarterly averages of monthly data, are from 1948:Q1–2004:Q1. Employment, unemployment, and short-term unemployment are measured by the BLS from the CPS, and the remaining variables are constructed as described in the paper, with $f_t^e = af_t$. The unemployment rate is shown for comparison.

[8] These data can be downloaded from the NBER web site: www.nber.org/data/cps_basic.html.

Figure 5

Empirical and Theoretical Job-to-Job Transition Rate



NOTE: Data, quarterly averages of monthly data, are from 1994:Q1–2004:Q1. The empirical job-to-job transition rate is computed from public use CPS micro data as the fraction of employed workers who have a "new employer" rather than the "same employer as last month." The theoretical job-to-job transition rate is computed using equation (12) with $a = 0.15$.

analyze only one cyclical downturn using this measure.

With these caveats, Figure 5 shows the empirical behavior of this measure of the job-to-job transition rate, with the prediction of the theoretical model ($a = 0.15$) included for comparison. The fact that the levels are approximately correct is due to a judicious choice of the relative effectiveness of on-the-job search a. But the underlying data on job-finding and separation rates drive the fluctuations in the theoretical series. Although the theory predicts a moderate increase in the job-to-job transition rate during the second half of the 1990s while the data show little trend, the model is reasonably good at explaining the decline in job-to-job transitions from 2001 to 2003.

Another measure of the job-to-job transition rate uses the public-use micro data in the March supplement to the CPS (www.nber.org/data/cps.html). Since 1976, the March supplement has asked workers how many employers they had in the previous year. A coarse measure of the number of job-to-job transitions is $\sum_{i=1}^{\infty} n_i(i-1)$, where n_i is the number of workers reporting i employers. I multiply this by 52 and divide by the total number of weeks worked in the previous year to obtain the job-to-job transition rate per full year of work. Of course, many workers have a spell of unemployment in between employers, so this should be an upper bound on the job-to-job transition rate. To obtain a lower bound, I look only at people who reported working for 52 weeks. An intermediate estimate follows Blanchard and Diamond (1990) and examines a question about the number of spells of job search. I assume that a worker who reports i employers and j spells of job search had $\max\{i-j-1,0\}$ job-to-job transitions during the previous year.

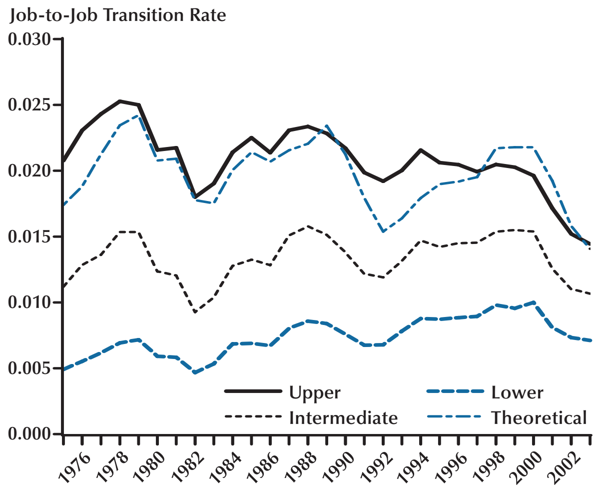
The main advantage to this measure is that the necessary questions have been asked every year since 1976.[9] But there are three significant disadvantages. The data are available only on an annual basis, a lower frequency than is ideal for business cycle analysis. The questions are retrospective, so respondents may forget some job-to-job transitions. And the number of job-to-job transitions is capped at three per year, which understates the true job-to-job transition rate. Figure 6 shows the upper, intermediate, and lower measures of the job-to-job transition rate based on March CPS data. The most striking finding is that even the upper bound lies significantly below the estimate using the monthly CPS (Figure 5). But despite this difference in the level, the March CPS data confirm that the job-to-job transition rate is procyclical. In fact, Figure 6 shows that the theoretical prediction of the job-to-job transition rate with a low value of the relative efficiency of on-the-job search ($a = 0.05$) closely tracks the upper-bound empirical estimate.

A final method of estimating the job-to-job transition rate comes from the Job Openings and Labor Turnover Survey (JOLTS; www.bls.gov/jlt). Since December 2000, this BLS survey has asked business establishments how many workers they have added to their payrolls during the previous

[9] The question in the year-t CPS asks about employers in year $t-1$, so I have data from 1975 to 2003.

Figure 6

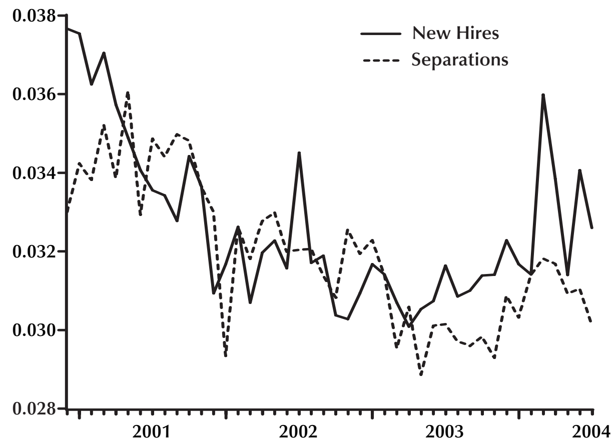
Empirical and Theoretical Job-to-Job Transition Rate, 1975-2003



NOTE: The upper, intermediate, and lower empirical job-to-job transition rates are computed from the public use files of the March CPS, as described in the text. The theoretical job-to-job transition rate is computed using equation (12) with $a = 0.05$.

Figure 7

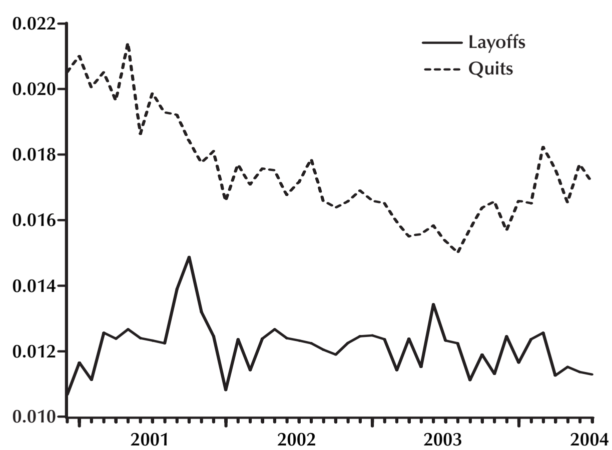
New Hires and Total Separations as a Percent of Employment



NOTE: Data are from December 2000 through July 2004 (seasonally adjusted) and are constructed by the BLS as part of the JOLTS program.

Figure 8

Layoffs and Quits as a Percent of Employment



NOTE: Data are from December 2000 through July 2004 (seasonally adjusted) and are constructed by the BLS as part of the JOLTS program.

month, how many workers left their payrolls, and whether those workers were laid off or quit. Figure 7 suggests that both new hires and separations fell as the United States labor market remained weak in 2002 and 2003. More tellingly, Figure 8 shows only a brief small spike in layoffs just after the terrorist attacks in September 2001, while the number of instances in which a worker quit her job fell steadily during this period.[10] For present purposes, it is notable that both new hires and "quits" fell by approximately 0.6 percentage points between 2001 and 2003. This finding is most easily explained by a decline in voluntary job-to-job transitions of the same magnitude, which reduced both new hires and quits.

In summary, if downturns were periods with high separation rates and normal job-finding rates, the model described in this paper would predict

[10] Akerlof, Rose, and Yellen (1988) discuss an old BLS survey of turnover in manufacturing establishments. Their Figure 1 shows that quits were strongly procyclical, whereas their Figure 2 shows brief spikes in layoffs at the start of a downturn. Unfortunately, the BLS discontinued this survey in 1982.

an increase in job-to-job transitions during downturns, and in particular from 2001 to 2003. This increase would occur both because of an increase in the number of workers who suffer a separation but manage to find a new job and because the increase in separations would reduce the age of matches and hence their quality, causing more voluntary job-to-job transitions. The finding that job-to-job transitions typically fall during downturns is consistent with the evidence that workers find it harder to obtain a job during downturns and do not experience a large increase in their separation rate.

CONCLUSIONS

This paper argues that business cycle fluctuations in unemployment are primarily a consequence of changes in the probability that an unemployed worker finds a job within a month—the job-finding probability. Changes in the separation rate do not explain any of the observed unemployment fluctuations during the past two decades. A simple model of on-the-job search suggests that in such an environment, job-to-job transitions should be procyclical, consistent with recent evidence from JOLTS and the monthly CPS and historical evidence from the March CPS. In contrast, if the job-finding rate were acyclical and the separation rate countercyclical, the simple model would predict that the job-to-job transition rate would be countercyclical.

The interesting follow-up question is why there are times when it is so hard to find a job. This paper does not provide an answer, but some recent research by Kraus and Lubik (2004) and Nagypál (2004b) suggests that job-to-job transitions may be an important element of the answer.[11] Kraus and Lubik (2004) make two important assumptions. First, they assume that on-the-job search intensity is highly elastic, so a small change in labor market conditions leads to a large change in the amount of search by employed workers. Second, they work in a multisector model in which the output of each sector is complementary to the production of a final consumption good. When an employed worker switches sectors, the marginal revenue product of labor in the originating sector increases, giving an incentive for firms in that sector to create new vacancies. This creates something analogous to a vacancy chain (Akerlof, Rose, and Yellen, 1988). In this environment, a small positive productivity shock induces much more search by employed workers. As those workers start to find jobs in more-productive sectors, the low-productivity sectors create additional job openings, which draw workers out of unemployment. This generates large fluctuations in the vacancy-unemployment ratio. Although this research appears promising, my analysis in this paper assumed that employed workers' search intensity is inelastic and generated approximately correct fluctuations in the job-to-job transition rate. This result suggests that if employed workers' search intensity is highly procyclical, the model will generate too-large fluctuations in the job-to-job transition rate.

Nagypál (2004b) also gives job-to-job transitions a prominent role in her analysis. She proposes that to understand the cyclicality of vacancies, one has to understand whether firms prefer to hire employed or unemployed workers. She argues that firms prefer employed workers because, when an employed worker is willing to take a job offer, this is a strong signal that the worker likes the job. In contrast, unemployed workers will accept any job but continue to search for a better one. If turnover is costly and a worker cannot be forced to bear the full cost of her decision to leave a job, firms will then prefer to hire employed workers. And because employed workers are a relatively large portion of the searching population during booms, this encourages firms to create vacancies. Whether or not either of these explanations is ultimately proved correct, it seems likely that a satisfactory model of the job-finding rate will explain why it is harder both for unemployed and employed workers to find jobs during downturns.

[11] Some other recent research that attempts to explain fluctuations in the job-finding rate focuses on wage rigidities arising either from social norms or asymmetric information. See Hall (2004), Kennan (2004), and Shimer and Wright (2004).

REFERENCES

Abraham, Katharine G. and Shimer, Robert. "Changes in Unemployment Duration and Labor-Force Attachment," in Alan Krueger and Robert Solow, eds., *The Roaring Nineties: Can Full Employment Be Sustained?*" New York: Russell Sage Foundation and Century Foundation, 2001, pp. 367-420.

Akerlof, George A.; Rose, Andrew K. and Yellen, Janet L. "Job Switching and Job Satisfaction in the U.S. Labor Market." *Brookings Papers on Economic Activity*, 1988, (2), pp. 495-582.

Barlevy, Gadi. "The Sullying Effect of Recessions." *Review of Economic Studies*, January 2002, *69*(1), pp. 65-96.

Blanchard, Olivier Jean and Diamond, Peter. "The Cyclical Behavior of the Gross Flows of U.S. Workers." *Brookings Papers on Economic Activity*, 1990, (2), pp. 85-143.

Boeri, Tito. "Is Job Turnover Countercyclical?" *Journal of Labor Economics*, October 1996, *14*(4), pp. 603-25.

Burdett, Kenneth. "A Theory of Employee Job Search and Quit Rates." *American Economic Review*, March 1978, *68*(1), pp. 212-20.

Burdett, Kenneth and Mortensen, Dale T. "Wage Differentials, Employer Size, and Unemployment." *International Economic Review*, May 1998, *39*(2), pp. 257-73.

Caballero, Ricardo J. and Hammour, Mohamad L. "The Cleansing Effect of Recessions." *American Economic Review*, December 1994, *84*(5), pp. 1350-68.

Cole, Harold L. and Rogerson, Richard. "Can the Mortensen-Pissarides Matching Model Match the Business-Cycle Facts?" *International Economic Review*, November 1999, *40*(4), pp. 933-59.

Davis, Steven J. and Haltiwanger, John. "Gross Job Creation and Destruction: Microeconomic Evidence and Macroeconomic Implications" in Olivier Jean Blanchard and Stanley Fischer, eds., *NBER Macroeconomics Annual 1990.* Cambridge, MA: MIT Press, 1990, pp. 123-68.

Davis, Steven J. and Haltiwanger, John C. "Gross Job Creation, Gross Job Destruction, and Employment Reallocation." *Quarterly Journal of Economics*, August 1992, *107*(3), pp. 818-63.

Davis, Steven J.; Haltiwanger, John C. and Schuh, Scott. *Job Creation and Destruction*. Cambridge, MA: MIT Press, 1996.

den Haan, Wouter; Ramey, Garey and Watson, Joel. "Job Destruction and Propagation of Shocks." *American Economic Review*, June 2000, *90*(3), pp. 482-98.

Fallick, Bruce and Fleischman, Charles. "Employer-to-Employer Flows in the U.S. Labor Market: The Complete Picture of Gross Worker Flows." Finance and Economics Discussion Series Working Paper 2004-34, Board of Governors of the Federal Reserve System, 2004.

Foote, Christopher. "Trend Employment Growth and the Bunching of Job Creation and Destruction." *Quarterly Economic Review*, August 1998, *113*(3), pp. 809-34.

Hall, Robert E. "Lost Jobs." *Brookings Papers on Economic Activity*, 1995, (1), pp. 221-56.

Hall, Robert E. "The Labor Market Is Key To Understanding the Business Cycle." Unpublished manuscript, Stanford University, September 23, 2004.

Hall, Robert E. "Employment Efficiency and Sticky Wages: Evidence from Flows in the Labor Market." *Review of Economics and Statistics*, 2005a (forthcoming).

Hall, Robert E. "Employment Fluctuations with Equilibrium Wage Stickiness." *American Economic Review*, 2005b, *95*(1), pp. 50-65.

Kennan, John. "Private Information, Wage Bargaining and Employment Fluctuations." Unpublished manuscript, University of Wisconsin–Madison, April 2004.

Kraus, Michael and Lubik, Thomas. "On-the-Job Search and the Cyclical Dynamics of the Labor Market." Unpublished manuscript, University of Tilburg, 2004.

Mortensen, Dale T. and Pissarides, Christopher A. "Job Creation and Job Destruction in the Theory of Unemployment." *Review of Economic Studies*, July 1994, *61*(3), pp. 397-415.

Nagypál, Éva. "Amplification of Productivity Shocks: Why Vacancies Don't Like To Hire the Unemployed?" Unpublished manuscript, Northwestern University, 2004a.

Nagypál, Éva. "Worker Reallocation over the Business Cycle: The Importance of Job-to-Job Transitions." Unpublished manuscript, Northwestern University, 2004b.

Pissarides, Christopher A. "Search Unemployment with On-the-Job Search." *Review of Economic Studies*, July 1994, *61*(3), pp. 457-75.

Polivka, Anne E. and Miller, Stephen M. "The CPS after the Redesign," in John Haltiwanger, Marilyn Manser, and Robert Topel, eds., *Labor Statistics Measurement Issues*. Chicago: University of Chicago Press, pp. 249-89.

Pries, Michael J. "Persistence of Employment Fluctuations: A Model of Recurring Job Loss." *Review of Economic Studies*, January 2004, *71*(1), pp. 193-215.

Ramey, Garey and Watson, Joel. "Contractual Fragility, Job Destruction, and Business Cycles." *Quarterly Economic Review*, August 1997, *112*(3), pp. 873-911.

Shimer, Robert. "The Cyclical Behavior of Equilibrium Unemployment and Vacancies." *American Economic Review*, 2005a, *95*(1), pp. 25-49.

Shimer, Robert. "Reassessing the Ins and Outs of Unemployment." Unpublished manuscript, University of Chicago, 2005b.

Shimer, Robert and Wright, Randall. "Competitive Search Equilibrium with Asymmetric Information." Unpublished manuscript, University of Chicago, August 26, 2004.

Solon, Gary. "Effects of Rotation Group Bias on Estimation of Unemployment." *Journal of Business and Economic Statistics*, January 1986, *4*(1), pp. 105-09.

APPENDIX

MEASUREMENT OF SHORT-TERM UNEMPLOYMENT

To measure short-term unemployment, I rely on workers' self-reported duration of an in-progress unemployment spell. Unfortunately, the CPS instrument was redesigned in January 1994, changing how the unemployment duration question was asked (Abraham and Shimer, 2001).[12] Recall that the CPS is a rotating panel. Each household is in the CPS for four consecutive months (rotation groups 1 to 4), out for eight months, and then in again for four more months (rotation groups 5 to 8). This means that in any month, approximately three-quarters of the households in the survey were also interviewed in the previous month.

Until January 1994, unemployed workers in all eight rotation groups were asked how long they had been unemployed. But since then, the CPS has not asked a worker who is unemployed in consecutive months the duration of her unemployment spell in the second month. Instead, the BLS calculates unemployment duration in the second month as the sum of unemployment duration in the first month plus the intervening number of weeks. Thus, prior to 1994, the CPS measure of short-term unemployment

[12] See Polivka and Miller (1998) for a thorough analysis of the redesign of the CPS instrument.

should capture the total number of unemployed workers who were employed at any point during the preceding month, whereas after the redesign, short-term unemployment captures only workers who transition from employment at one survey date to unemployment at the next survey date.[13]

There is no theoretical reason to prefer one measure to the other, because they simply measure different objects. But the method I use to measure the job-finding and separation rates in the second section of this paper ("Measuring the Job-Finding and Separation Rates") relies on the pre-1994 measure of short-term unemployment. In any case, the goal of this paper is to obtain a consistent time series for the job-finding rate. To obtain such a time series, note that one would expect that the redesign of the CPS instrument would not affect measured unemployment duration in rotation groups 1 and 5, the "incoming rotation groups," because these workers are always asked their unemployment duration, but would reduce the measured short-term unemployment rate in the remaining six rotation groups.

To see this empirically, I use CPS micro data from January 1976 to March 2004 (www.nber.org/data/cps_basic.html) to measure short-term unemployment. In an average month from January 1976 to January 1994, short-term unemployment accounted for 41.6 percent of total unemployment in the full CPS and 41.7 percent in the incoming rotation groups, an insignificant difference. From February 1994 to March 2004, however, short-term unemployment accounted for 38.6 percent of unemployment in the full sample but 44.6 percent in the incoming rotation groups, an economically and statistically significant difference. Put differently, the short-term unemployment rate in the full CPS fell discontinuously between January and February 1994, while it remained roughly constant in the incoming rotation groups.

In this paper I use short-term unemployment from the full sample from January 1948 to January 1994 and then use only the incoming rotation groups in the later period. More precisely, I multiply the number of unemployed workers in the full CPS sample by the fraction of short-term unemployed among unemployed workers in the incoming rotation groups.[14] This avoids any discontinuity associated with the redesign of the CPS.[15]

---

[13] The post-1994 methodology also prevents respondents from erroneously reporting short unemployment duration month after month.

[14] This approach circumvents another problem. From 1976 to 2004, the unemployment rate in the first rotation group averaged 0.4 percentage points more than in the full sample. See Solon (1986) for an analysis of rotation-group biases in the CPS.

[15] I have also tried multiplying the standard series for short-term unemployment by a constant, 1.1, after January 1994, which delivers very similar results.

Commentary

Randall Wright

As Robert Shimer (2005) emphasizes, modern theories of the labor market recognize the conceptual value of decomposing unemployment fluctuations into hires and separations. Thus, we have

$$u_{t+1} = u_t(1-h_t) + (1-u_t)s_t,$$

where u_t denotes unemployment at date t, h_t denotes the hiring rate at t, and s_t denotes the separation rate at t. Or, if we want to look at changes,

$$\Delta u_t = u_{t+1} - u_t = (1-u_t)s_t - u_t h_t.$$

If, for example, h_t and s_t are constant with respect to t, then starting at any initial u_0,

$$u_t \rightarrow u^* = \frac{s}{s+h}.$$

This kind of discussion is now standard fare in graduate and good undergraduate courses.[1]

Here is a question: Given $\Delta u_t > 0$, as might happen during the downturn of a typical recession, is it because separations are high or because hires are low, and does it matter? The version of the paper that I read contained some examples attempting to illustrate just why this matters, and I want to discuss them briefly. As the first example, consider the sectoral shift hypothesis, which is the idea that recessions are best thought of as declines in some sectors. This decline, it is suggested, will show up as an increase in the separation rates, $\Delta_s > 0$, with little change in the hiring rates, $\Delta h = 0$. But why? We know h can be big even in recessions. Declining firms could keep s the same, reduce h, and downsize through attrition. So without further elaboration, and I don't see what this would be, the decomposition of changes in u into (i) the component due to changes in h and (ii) the component due to changes in s does not strike me as somehow being the key to evaluating the sectoral shift hypothesis.

A second example in the paper concerned the literature on firing costs. According to the discussion, this literature thinks of recessions as times when $\Delta s = 0$ and $\Delta h < 0$, which is in some sense the opposite of the sectoral shift hypothesis previously mentioned. Hence, knowing from the data whether recessions are times when $\Delta s = 0$ and $\Delta h < 0$ or times when $\Delta s > 0$ and $\Delta h = 0$ is the key to distinguishing between these stories. But, wait a minute: Does anyone actually think firing-cost models and sectoral-shift models are competing world views? Why aren't they two aspects of one story? At some level this reflects a confusion between impulse and propagation mechanisms: Sectoral shifts could be the underlying cause of economic downturns, while firing costs could have effects on the dynamics. Note that I am not trying to champion this position here; I simply want to ask what the fuss is about.

A third example concerns the Keynesian litera-

[1] Shimer tries to make a distinction between this approach and an earlier literature that decomposes changes in u_t into changes in *incidence* and *duration* of unemployment. This is tenuous because the incidence probability is simply s and expected duration is simply $1/h$. Perhaps there are situations (e.g., with time aggregation) where there is a relevant distinction.

Randall Wright is a professor of economics at the University of Pennsylvania and a Central Bank Institute Scholar at the Federal Reserve Bank of Cleveland.

Federal Reserve Bank of St. Louis *Review*, July/August 2005, *87*(4), pp. 509-12.

ture. The idea seems to be that if there are nominal rigidities in wages, but new hires can get around these rigidities, then in fact recessions are actually times when firms really want to hire new workers. This strikes me as a bit silly, as it seems to imply that the Keynesian model predicts that we can avoid recessions (caused by sticky nominal wages) by having firms swap workers. Whatever. I don't think Shimer was ever very happy with these examples, and that is why they do not appear in the final version. So why am I discussing them? Well, first, maybe the points contained in this discussion are at least part of the reason why they were left on the cutting room floor; and second, it seems to me that we do need to think more about why the decomposition of changes in u into the part due to changes in s and the part due to changes in h is interesting.

Perhaps it is interesting for its own sake. Fine. I am quite prepared to take this as given for now. The real motivation for the paper is "to document the cyclicality of the hiring and separation rates in the United States for 1948-2004." I think this is a good idea; at least, it's "something to do." Let it be said that this is not as easy as one might think. For example, a big effort is made to take into account composition effects due to heterogeneous agents and movements in and out of the labor force, as I will discuss here. Also, the results do at least provide a clear cut answer to the previously mentioned question: The finding is that there are substantial fluctuations in h, and much less in s, and indeed fluctuations in s are even described at one point as "acyclical."[2]

Another of Shimer's findings—"perhaps the most surprising"—is the following: "Whatever forces make it harder for an unemployed worker to find a job during a recession also seem to make it harder for an employed worker to find a better job." Is this really *surprising*? Perhaps not, but it does suggest that it is worth thinking about the cyclical behavior of hiring rates seriously. For this, one needs a good measure of h, and at the end of the day this is what the exercise is all about. Given this, it seems to me that Shimer does an admirable job coming up with a new and improved measure. Again, heterogeneity in h^i across agents is relevant. In particular, with homogenous agents, the different possible measures compared in the paper are equivalent, and hence it is only with heterogeneity that Shimer's measure is therefore either new or improved. But there should be little doubt that heterogeneity may be important in this context, and so considering this new measure of h seems useful.

Let me move on to some more detailed discussion of the actual exercise. Suppose we define *short-term unemployment* by $u^s_{t+1} = (1 - u_t)s_t$; then we have

$$u_{t+1} = u_t\left(1 - h_t\right) + u^s_{t+1},$$

at least under the hypothesis that the labor force is constant (see forthcoming description). The advantage of this formulation is that we have a direct measure of u^s_{t+1} in the data. Hence, we can construct

$$h_t = 1 - \frac{u_{t+1} - u^s_{t+1}}{u_t}$$

as an empirical measure of hiring. Notice that, even if we have heterogeneity in hiring rates, $h_t = E_i h^i_t$. A few facts: $E_t h_t = 0.44$; and (after filtering), $cor(h_t,u_t) = -0.94$, $sd(h_t)$–0.12, and $sd(u_t) = 0.20$.[3]

Now we get down to some serious issues. First, compositional effects. This issue is simple: Do hiring rates really *change* during recessions, or are there just more low-h people in the unemployment pool during recessions? Shimer dismisses the importance of compositional effects, but not at all casually. He does a good job of trying to address the problem, but it is a problem that as a matter of principle can never be resolved to full satisfaction. For example, suppose we have two types of workers, type L and type H, where $h^L < h^H$. Now suppose that, for any number of reasons that are easy enough to imagine, more type L workers lose their jobs in a recession. Shimer's approach is to divide the workers in his sample

[2] Saying they are "acyclical" seems to be a slight exaggeration, but this perhaps is a quibble about semantics.

[3] As is typical in the literature, Shimer goes on to emphasize that "$sd(u_t)$ is 70 percent larger than and $sh(u_t)$"—which seems to add little to our knowledge once we have been told that $sd(h_t) = 0.12$ and $sd(u_t) = 0.20$.

into groups by race, sex, and so on. But although this is useful, it does *not* completely resolve the potential problem.

For example, suppose that my two types are *lazy* workers with low h^L and *not-so-lazy* workers with high h^H. What observable characteristic is Shimer going to say proxies for laziness—race or sex? Obviously this is not going to be a viable approach. Of course, what I am saying here is obvious: There is no way to *reall*y control for unobserved heterogeneity. So while I am sympathetic to what Shimer has done, and perhaps it is the best that could be done, it is simply not a definitive result that heterogeneity is not what is driving the behavior of his empirical measure of h. I shall not beat this to death; but at the same it does deserve mention.

Shimer goes on to worry about changes in the labor force. Suppose we drop the maintained hypothesis that the labor force is fixed. Letting x_t denote exit from the labor force, we have the following new version of our law of motion for u_t,

$$u_{t+1} = u_t\left(1 - h_t - x_t\right) + u_{t+1}^s,$$

and hence we have

$$h_t + x_t = 1 - \frac{u_{t+1} - u_{t+1}^s}{u_t}.$$

Therefore, his empirical variable h_t is really measuring the exit rate from u_t, either into employment or out of the labor force, and not the hiring rate, per se.

Shimer argues that empirically this is not a big deal. Fine, but there is a sense in which a whole other issue is raised. At least for some purposes, we might prefer to think about the world in terms of a two-state model, where e workers are employed and $1 - e$ are not employed; and let's not worry about decomposing the latter group into those in and those out of the labor force. Because, in reality, virtually everyone is "in the labor force" in the sense that they would be willing to take some job (the right job) if it came along, even if the official data do not recognize this. Of course, some people who are not working are more actively searching for work, or more willing to accept work, than others; this is a matter of degree, and it is obviously quite arbitrary to define some criterion by which we label some in and others out of the labor force. One may say this is a side issue in terms of the focus of the current project, but given we are engaging in a careful measurement exercise, it is not illegitimate to ask what it is we ought to be trying to measure.

To move on, in addition to constructing a measure of the hiring rate, h, the paper also considers the separation rate, s, for the purpose of comparison. This is somewhat tricky because of time-aggregation issues. Shimer takes this seriously, but it is a slippery slope. He worries in particular about false job-to-job transitions recorded in the data, because a worker may have lost a job and found a new one between surveys. Sure. But what about the other side of the coin? Obviously many people have jobs, line up new ones before either leaving or losing the old jobs, but spend a little time between jobs doing things like moving, collecting unemployment insurance, chilling out, or whatever. These may look like transitions from e to u to e, but for many purposes it seems better to think of them as "really" more like job-to-job transitions. Can we tell from the data? Does it matter? Is there any way to resolve the issue satisfactorily? These seem like good questions.

By the way, there is also a model in the paper—a model with on-the-job search, of course, since how else could one expect to discuss the large number of job-to-job transitions in the data. Shimer assumes for simplicity that the distribution of job offers is time invariant. Given this, after some routine on-the-job-search algebra we get some nice results. He uses the model to come up with measures of how many workers switch jobs each month *involuntarily* (i.e., with an intervening spell of unemployment) and *voluntarily* (i.e., without same). I am not sure these words constitute the best choice of language, especially given the time-aggregation issues raised in the previous paragraph, but this is his choice. Another issue is that the numbers he comes up with are sensitive to assumptions about arrival rates of offers for employed and unemployed workers. What should we do about pinning down these arrival rates? Perhaps one can calibrate to match the observations on labor market flows in, say, Fallick and

Fleischman (2001). Or one can look to empirical estimates of the Burdett and Mortensen (1998) model. Or one can make them up.

Let me try to wrap up. First, I want to say that this is agreeable work. As Finn Kydland (or is it Ed Prescott?) often says, "we should all be in favor of good measurement." The finding is this: It is changes in *h* and not changes in *s* that drive fluctuations in *u*. Shimer's conclusion is that the "received wisdom" about "job destruction rising dramatically during recessions" may need to be reassessed. I found the argument compelling. There are outstanding issues. Does unobserved heterogeneity mean we can never really know if *h* varies over the cycle—or if it only seems to—due to compositional effects? Do we want to think about the labor market in terms of two states, say, employed and not employed, or three? What should we make of the time-aggregation problems, and how can we best correct for potential false job-to-job transitions as well as false job-to-unemployment-to-job transitions? As usual, reading a paper by Shimer made me think about many interesting things, including several that are not completely resolved in the paper. And although I see how this could be taken either way, in this case it is meant as a compliment.

REFERENCES

Burdett, Kenneth and Mortensen, Dale T. "Wage Differentials, Employment Size, and Unemployment." *International Economic Review*, May 1998, *39*(2), pp. 257-73.

Fallick, Bruce C. and Fleischman, Charles A. "The Importance of Employer-to-Employer Flows in the U.S. Labor Market." Finance and Economics Discussion Series 2001-18, Board of Governors of the Federal Reserve System, April 2001.

Shimer, Robert. "Employer-to-Employer Transitions and the Business Cycles." Federal Reserve Bank of St. Louis *Review*, July/August 2005, *87*(4), pp. 493-507.

Reexamining the Monetarist Critique of Interest Rate Rules

Robert G. King and Mau-Ting Lin

Monetarist economists argued long ago that central bank interest rate rules exacerbate macro economic fluctuations, essentially by not allowing the interest rate to respond promptly to shifts in the supply and demand for loans. To support this critique, they pointed to the procyclicality of the money stock. Yet, when there are real shocks and a real business cycle, modern macroeconomic models imply that some procyclicality of money is desirable, to stabilize the price level. A simple interest rate rule illustrates that the monetarist critique can be valid within this model, since the rule exacerbates the response of real activity to real shocks. Other interest rate rules instead limit the macro economy's response to real shocks. But, while these interest rate rules have diverse effects on real activity, there is an important common implication: By smoothing the nominal interest rate in the short run, the rules all lead to increases in the longer-run variability in inflation and nominal interest rates.

Federal Reserve Bank of St. Louis *Review*, July/August 2005, *87*(4), pp. 513-30.

1 INTRODUCTION

Once upon a time, the nature of the short-term nominal interest rate was a central dispute among macroeconomists. Keynesian economists stressed that it was the opportunity cost of holding money. Monetarist economists stressed that the nominal rate was a central part of an intertemporal price, with the real interest rate the equating market for loan supply and demand.

Such divergent perspectives led these groups of macroeconomists to subscribe to different policies for the management of interest rates. Viewing the demand for money as fluctuating substantially over time and viewing the short-term nominal interest rate as principally affected by monetary factors, Keynesian macroeconomists argued for holding the interest rate fixed as economic activity fluctuated or, at least, varying it gradually over time. Viewing loan supply and demand as subject to important real shocks, monetarist economists argued for allowing interest rates to fluctuate more widely, while seeking to control a monetary aggregate for stabilization purposes. More specifically, Brunner (1978), Friedman (1968, 1985), and particularly Poole (1978) argued that the Federal Reserve System's unwillingness to change interest rates over time exacerbated the business cycle, leading the central bank to make the money supply procyclical.[1]

In recent years, macroeconomic analysis has increasingly used small, fully articulated quantitative macroeconomic models to study the effects of alternative monetary policies. In the terminology of King and Wolman (1996), these are the "St. Louis models of the 21st Century," capable of exploring alternative policies in a manner

[1] For example, in an interview in this *Review*, Brunner (1978) argues "apparently uncontrollable money growth… essentially results from the central bank's unwillingness, or political inability, to adjust the interest rate…to the realities of the market place."

Robert G. King is a professor of economics at Boston University, a research associate at the National Bureau of Economic Research, and an editor for the *Journal of Monetary Economics*. Mau-Ting Lin is an assistant professor of economics at the National University of Singapore.

consistent with the recommendations of Lucas (1976).[2] Increasingly, these models are studied with a focus on interest rate rules for monetary policy, particularly variants of the rule put forward by Taylor (1993). In this article, we return to the concerns of Brunner, Friedman, and Poole, studying the effect of some alternative interest rate rules within a modern quantitative macroeconomic model. We are specifically interested in whether monetarist concerns about the exacerbation of the business cycle carry through to modern model economies under alternative interest rate rules related to Taylor's specification. Therefore, in our analysis, we consider three interest rate rules and we explore how each affects the dynamic response of the macroeconomy to two real shocks: shifts in government purchases and productivity. First, we consider an "inflation only" variant of the Taylor rule, in which there is no output gap response. Second, we consider the original Taylor specification. Third, we consider a more dynamically complicated interest rate rule estimated by Orphanides and Wieland (1998), which includes a lagged nominal interest rate term.

To consider how these rules alter the behavior of economic activity, we exploit the fact that there is a "neutral" solution for real activity—the solution that would obtain under flexible prices—that can be brought about by the monetary authority if it fully stabilizes the price level. The neutral real activity solution can be brought about by an interest rate rule that "tracks the natural rate of interest." We use the neutral outcomes as a benchmark for subsequent analysis. The analysis of neutral outcomes also highlights the fact that the procyclicality of money does not, on its own, rationalize the concerns of Brunner, Friedman, and Poole. Since the neutral output solution requires that the price level be stabilized, while real activity fluctuates, money must move positively with real activity. In this regard, our model accords with Friedman's (1969, p. 46) observation that a stable price level requires a specific trend growth rate of money when there is growth in productivity and labor, but we apply this reasoning to output responses over the course of a real business cycle.

Using the "inflation only" interest rate rule, we find that this policy exacerbates economic fluctuations for the reasons suggested by Brunner, Friedman, and Poole. Since nominal and real rates should rise immediately in response to both disturbances in the neutral solution, but cannot according to the policy rule, there must be an additional stimulative increase in money. Thus, there is a temporary increase in output relative to the neutral paths relevant for each disturbance. The key to these exacerbation results is that the "inflation only" interest rate rule does not accommodate shifts in the neutral real interest rate.

The other interest rate rules that we consider, the specifications of Taylor (1993) and Orphanides and Wieland (1998), also do not accommodate shifts in the neutral real rate, as shocks to government purchases or productivity take place. However, because there are responses to an output gap measure in these rules, their implications are more complicated. We find that these rules restrict—rather than exacerbate—cyclical variations in output that arise from real shocks. In effect, when there is a rise in real economic activity from our demand shock (government purchases) or supply shock (productivity), these rules call for a smaller degree of monetary accommodation than would occur under neutral policy. Hence, increases in government purchases and productivity both bring about declines in the price level for several quarters, which are associated with temporary declines in output relative to the neutral level. Strikingly, given the empirical analysis of Galí (1999), both our Taylor and Orphanides-Wieland specifications imply that there is a small *absolute* decline in output when there is a positive productivity shock, which translates into an important decline in labor input.[3]

While the interest rate rules that we study have

[2] Kimball (1995) calls similar models "neomonetarist." Goodfriend and King (1997) describe them as the result of a "new neoclassical syntheses."

[3] Working within a modern macroeconomic model, but with staggered price setting at a 4-quarter horizon, Dotsey (2002) finds that the Taylor rule does not involve a decline in labor input. Although there are a number of other differences between the models, one is tempted to conclude that the discrepancy between his results and ours suggests that the effect of interest rate rules depends importantly on the details of the price-setting structure.

diverse consequences for real activity, there is an important common implication: By smoothing the nominal interest rate in the short-run, the rules all lead to a substantial increase in the variability of nominal interest rates. That is, the interest rate rules all lead to low-frequency variation in inflation, which is essentially neutral and thus fully reflected in the nominal rate. Such increased volatility in inflation and nominal interest rates is another standard monetarist concern about interest rate rules (Friedman, 1968, and Poole, 1978).

The organization of the paper is as follows. Section 2 describes the macroeconomic model that we employ in this paper. Section 3 discusses the response of the macroeconomy to real government purchase and real productivity shocks, under the assumption of "neutral policy," which stabilizes the path of the price level and produces outcomes that are equivalent to a real business cycle model. Section 4 considers how the dynamic responses differ under our three interest rate rules. Section 5 is a summary and conclusion.

2 A ST. LOUIS MODEL

The small quantitative macroeconomic model that we use is closely related to those in King and Watson (1996), King and Wolman (1996), and Yun (1996).[4]

2.1 Households

There are many identical households in the economy. Each values a composite consumption good, c, and leisure, l, as summarized by the expected utility objective

$$E_t \sum_{j=0}^{\infty} \beta^j u(c_{t+j}, l_{t+j}) . \tag{1}$$

The household can invest in various financial instruments, including government bonds at price $1/(1 + R_t)$, and in a diversified portfolio of claims to firms at price $\underline{v}_t$. It also makes investments, i_t, in real capital, which it rents to firms in the economy at rental rate, q_t. Its receives wage rate w_t for units of work n_t and receives lump-sum transfers or taxes from the government in amount T_t. Thus, its one-period budget constraint takes the form

$$\begin{aligned} & c_t + i_t + \frac{1}{1+R_t} b_{t+1} + \underline{v}_t e_{t+1} \\ & = \left[b_t \frac{P_{t-1}}{P_t} + (\underline{v}_t + \underline{z}_t) e_t + w_t n_t + q_t k_t + T_t \right]. \end{aligned} \tag{2}$$

In this expression, b_t and e_t are the quantities of bonds and equities held at the start of period t. The household faces three other constraints. First, capital evolves according to

$$k_{t+1} - k_t = h(\frac{i_t}{k_t}) k_t - \delta k_t . \tag{3}$$

In this expression, the rate of depreciation is δ and ϕ is an increasing, concave function ($h(0) = 0$), which allows for capital stock adjustment costs as in Hayashi (1982). Second, the household has a time constraint,

$$n_t + l_t \leq 1 . \tag{4}$$

Third, the household must hold an asset—fiat money—in sufficient quantity to pay for its consumption and investment expenditures, as well as tax payments to the government that are necessary to finance its real expenditures:

$$m_t = \frac{M_t}{P_t} = (c_t + i_t + g_t) . \tag{5}$$

We do not model the demand for money explicitly, simply assuming that the quantity equation holds.[5] In this regard, we depart from King and Wolman (1996), who introduced a shopping time technology to motivate money demand holding and derived detailed information on "transactions wedges" in an otherwise similar model. In the current analysis, we follow the strategy more

[4] The calibration and computation follow directly from King and Wolman (1996), with all parameters matching that study with one exception, which is that we assume a smaller extent of investment adjustment costs. The economy is linearized around a zero inflation steady state and solved as a linear rational expectations model. Replication materials for the analysis are available at http://people.bu/edu/rking/Research.htm.

[5] One way to think about this assumption is that it involves interest payments on the monetary aggregate, which is relevant for transactions purposes, but with a small spread so that the demand for money is determinate. Under this interpretation, it is appropriate to ignore revenue from money creation in the household budget constraint.

widely used in analysis of sticky price models, to stress that our results on policies are not dependent on a particular transactions technology and the implied money demand structure.

Given the problem just noted, we can cast the decision problem of households in dynamic programming form and derive efficiency conditions that restrict its optimal choices. First, individuals choose consumption and work efficiently:

(6) $$u_c(c_t, l_t) - \lambda_t = 0$$

(7) $$u_l(c_t, l_t) - \lambda_t w_t = 0.$$

In these expressions, λ_t is a Lagrange multiplier on (2) and w_t is a real wage rate. Second, individuals choose holdings of nominal and real bonds according to Fisherian principles,

(8) $$1 = \beta E_t[(1+R_t)\frac{\lambda_{t+1}}{\lambda_t}\frac{P_t}{P_{t+1}}]$$

(9) $$1 = \beta E_t[(1+r_t)\frac{\lambda_{t+1}}{\lambda_t}],$$

where r_t is a shadow real interest rate—which is determined so that the supply and demand for real bonds is zero—and R_t is the market-determined nominal interest rate.

2.2 Microgoods

Consumption, investment, and government aggregates are produced, using a standard Dixit-Stiglitz (1977) specification, from a continuum of goods on the unit interval. For example, the consumption aggregate is

$$c_t = [\int_0^1 [c_t(i)^{\frac{(\varepsilon-1)}{\varepsilon}} di]^{\frac{\varepsilon}{\varepsilon-1}}.$$

Cost minimization implies that the demand for the ith good takes the form

$$c_t(i) = (\frac{P_t(i)}{P_t})^{-\varepsilon} c_t,$$

where $P_t(i)$ is its nominal price and P_t is the price level, which Yun (1996) demonstrates takes the form

$$P_t = [\int_0^1 P_t(i)^{1-\varepsilon} di]^{\frac{1}{1-\varepsilon}}.$$

We assume that the same aggregator governs investment and government purchases, so that the demand for the ith product then takes the form

$$(\frac{P_t(i)}{P_t})^{-\varepsilon}(c_t + i_t + g_t),$$

where $c_t + i_t + g_t = d_t$ is aggregate demand.

2.3 Firms

As in many macroeconomic analyses with the "New Keynesian" elements surveyed in Rotemberg (1987), we assume that firms are monopolistic competitors facing exogenous opportunities for price adjustment in the Calvo (1983) manner. Each firm rents labor and capital from households, combining these factors to produce its output according to a constant returns-to-scale production function of the Cobb-Douglas form. Looking at firm i, its output is

(10) $$y_t(i) = a_t(k_t(i))^{1-\alpha}(n_t(i))^{\alpha},$$

where a_t is an aggregate productivity shifter, which is constant across firms. Given this specification, the firm's real marginal cost of producing output, ψ_t, is independent of its output level and equal to

(11) $$\psi_t = \frac{w_t^{\alpha} q_t^{1-\alpha}}{a_t[(1-\alpha)^{1-\alpha}\alpha^{\alpha}]}.$$

The absence of the firm index, i, reflects the fact that all firms have the same marginal cost, which varies with wage and rental rates. Firms set nominal prices, acting as monopolistic competitors and recognizing that they may be unable to adjust prices in future periods (with probability η). They have a forward-looking pricing rule of the form

(12)
$$\frac{P_{0t}}{P_t} = \frac{\varepsilon}{\varepsilon-1}\frac{\sum_{j=0}^{\infty}(\beta\eta)^j E_t\{(\lambda_{t+j}/\lambda_t)\cdot\psi_{t+j}\cdot d_{t+j}\}}{\sum_{j=0}^{\infty}(\beta\eta)^j E_t\{(\lambda_{t+j}/\lambda_t)\cdot(P_{t+j}/P_t)\cdot d_{t+j}\}}.$$

That is, the firm sets price as a markup over discounted measures of costs and demand, as discussed previously by King and Wolman (1996), Yun (1996), and others.

2.4 Aggregation

Since there is a continuum of firms on the unit interval (so that the law of large numbers applies) and since the probability of being unable to adjust in a given period is η, the stationary distribution of firms by age of price is $\omega_j = (1-\eta)\eta^j$ for $j = 0,1,2...$ Accordingly, as stressed by Yun, there is an aggregate production function that depends on linear aggregates of inputs used by firms with various vintages of prices, $y_t = a_t f(k_t, n_t)$,

with $k_t = \sum_{j=0}^{\infty} \omega_j k_{jt}$ and $n_t = \sum_{j=0}^{\infty} \omega_j n_{jt}$

that can be used for purposes such as extraction of a Solow residual. At the same time, the price level pertinent for the demand analysis previously noted takes a nonlinear form:

$$(13) \quad P_t = [\sum_{j=0}^{\infty} \omega_j (P_{jt})^{1-\varepsilon}]^{\frac{1}{1-\varepsilon}} = [(1-\eta)P_{0t}^{1-\varepsilon} + \eta P_t^{1-\varepsilon}]^{\frac{1}{1-\varepsilon}}.$$

National income identities require that

$$(14) \quad (c_t + i_t + g_t)\delta_t = y_t ,$$

with

$$(15) \quad \delta_t = \sum_{j=0}^{\infty} \omega_j (\frac{P_{jt}}{P_t})^{-\varepsilon}$$

being a measure of relative price distortions. Because we will assume that the economy is operating at a low (zero) inflation rate, variations in δ_t are locally unimportant.

2.5 Neutral Policy Rule

We close this model with an interest rate policy rule. Our benchmark is a "neutral" policy rule that involves strict inflation management. This interest rate rule takes the form

$$(16) \quad R_t = r_t^* + f[\log(P_t) - \log(\bar{P})].$$

In this expression, r_t^* is the level of the real interest rate—sometimes called the "natural rate" of interest—that would arise if output were continuously at its flexible price level. With a large value of f, this rule may alternatively be approximated by

$$R_t = r_t^*$$
$$\log(P_t) = \log(P_{t-1}) = \log(\bar{P}).$$

Under this policy rule, then, actual and expected inflation are always zero. As stressed by King and Wolman (1996) and Goodfriend and King (1997), the economy operates at its "natural rate" level, i.e., delivers the same time series of output and other real variables as would arise if prices were completely flexible.

2.6 Taylor Rules

The next two rules are based on the work of John Taylor (1993), with greater than one-for-one increases in the nominal interest rate when inflation exceeds a specified benchmark level (treated here as zero for expositional ease). We write Taylor rules as

$$R_t = r^* + \tau_1(\frac{1}{T_1}\sum_{j=1}^{T_1} \pi_{t-j}) + \tau_2(\log(y_{t-1}) - \log(\bar{y}_{t-1})).$$

That is, the nominal interest rate is adjusted in response to an average of recent inflation and in response to a measure of the output gap. Rather than specifying an economic model of the output gap, we let it be deviation of output from an average of past values,

$$\log(y_{t-1}) - \log(\bar{y}_{t-1}) = [\log(y_{t-1}) - \frac{1}{T_2}\sum_{j=1}^{T_2} \log(y_{t-j})].$$

We do not require that the lags in these expressions be the same; indeed, our Taylor rule cases involve responses to annual average inflation ($T_1 = 4$) and deviations of output from a 6-year moving average ($T_2 = 24$). We use this simple specification because we think it is a simple description of how standard capacity output measures behave over time.

It is important to stress that our formulation of this rule, like Taylor's original specification, does *not* involve the central bank "tracking the natural rate of interest." (That is, it is r^* not r_t^* that enters into the policy rule.) In this regard, it is potentially quite different from the neutral policy rule stated previously.

2.6.1 An "Inflation Only" Variant of the Taylor Rule. As a reference point, we begin with an interest rate rule in which the central bank reacts only to inflation and responds only to annual average inflation over the preceding year:

$$R_t = r^* + 1.5 \cdot \left(\tfrac{1}{4}\sum_{j=1}^{4}\pi_{t-j}\right). \tag{17}$$

In this setting, we use Taylor's 1.5 value for the coefficient τ_1. This coefficient value implies that there is a unique stationary rational expectations equilibrium.

2.6.2 A Standard Taylor Rule. Our version of the standard Taylor rule is

$$R_t = r^* + 1.5\left(\frac{1}{4}\sum_{j=1}^{4}\pi_{t-j}\right) + 0.5 \cdot [\log(y_{t-1}) - \log(\bar{y}_{t-1})]. \tag{18}$$

That is, in this case, we use both of Taylor's values, setting $\tau_1 = 1.5$ and $\tau_2 = 0.5$. The output gap measure is based on deviations from the 6-year (24-quarter) moving average, i.e.,

$$\log(\bar{y}_{t-1}) = \tfrac{1}{24}\sum_{j=1}^{24}\log(y_{t-j}).$$

These values also imply that there is a unique stationary rational expectations equilibrium.

2.7 An Estimated Rule

There has been much recent work on estimating interest rate rules for monetary policy. We employ a particular specification due to Orphanides and Wieland (1998), which takes the form

$$\begin{aligned} R_t = {} & r^* + 0.795(R_{t-1} - r^*) + 0.625 \cdot \left(\tfrac{1}{4}\sum_{j=1}^{4}\pi_{t-j}\right) \\ & + 1.17 \cdot [\log(y_t) - \log(\bar{y}_t)] \\ & - 0.97 \cdot [\log(y_{t-1}) - \log(\bar{y}_{t-1})], \end{aligned} \tag{19}$$

where we again take the output gap as a deviation from a 6-year moving average. In our model, these values also imply that there is a unique stationary rational expectations equilibrium.

3 NEUTRAL BENCHMARKS

Our starting point is the responses that obtain in the "core real business cycle model" within our sticky price model. There are two interpretations of these responses, which we use interchangeably, with one important exception discussed in detail hereafter.

The first interpretation is that prices are completely flexible, so that we are essentially looking at the workings of a real business cycle (RBC) model. This set of outcomes is not efficient, because there is monopoly power, but represents one definition of the natural rate of output (see Goodfriend and King, 1997, and Woodford, 2003).

The second interpretation is that monetary policy is conducted, as in section 2.5.1, to fully smooth the price level. Under this policy, real outcomes are those of the RBC solution, but nominal money must also be determined so that real balances satisfy the quantity equation at the specified price level. That is, activist monetary policy is required to smooth the price level in the presence of real shocks.

Hence, the two interpretations differ only on the path of money: It is not incorporated in the RBC solution (and could be anything), and it takes on particular values in the pegged price level case. Accordingly, our figures report the path of the money stock under the latter interpretation.

3.1 Response to Government Purchases

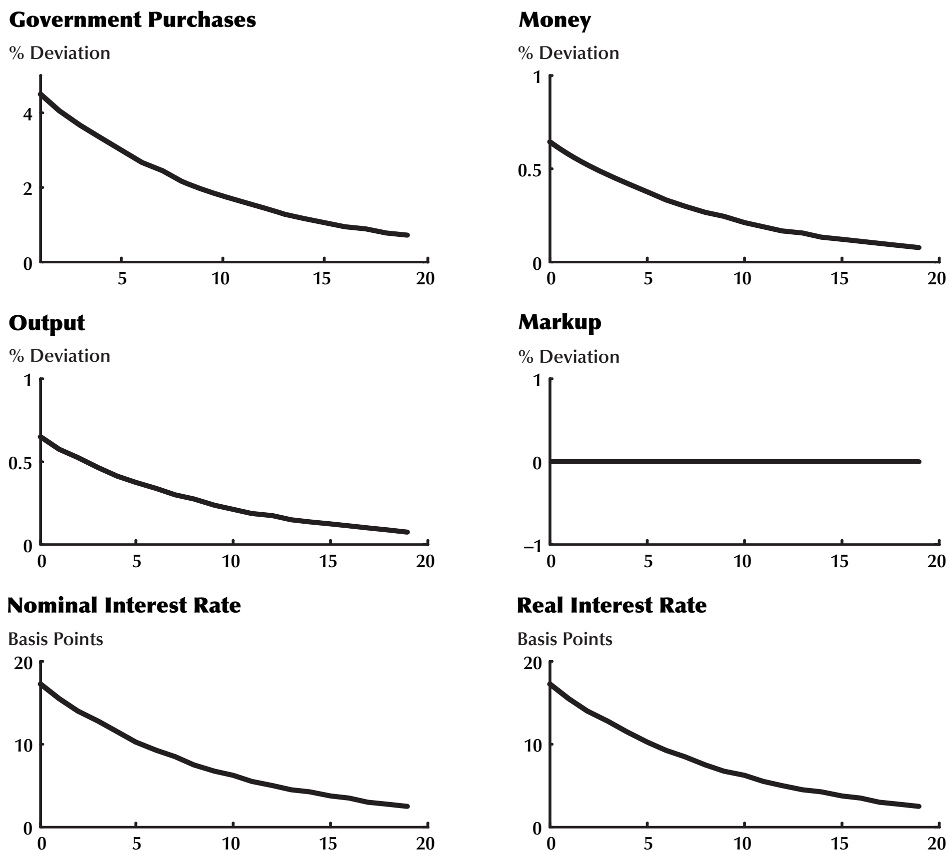
We consider a unit increase in government purchases under the assumption that the stationary share of government purchases is 20 percent ($g/y = 0.2$). We assume that this increase is persistent, but ultimately temporary, with government purchases being governed by $g_t = \rho g_{t-1} + e_{gt}$ with $\rho = 0.9$. The increase in government purchases is usefully thought of in two ways. First, it is a "demand shift," increasing the level of aggregate demand at a given real interest rate. Second, it is a "resource shift," amounting to a decrease in the output available for consumption and investment, without altering the level of the production function or its marginal products.[6]

The dynamic response to this shock is shown in Figure 1. The shock is a 5 percent movement in government purchases, which translates into

6 That is, in the terminology of Baxter and King (1993), we are considering the case of basic government purchases.

Figure 1

Effects of a Rise in Government Purchases Equal to 1 Percent of GDP Under Strict Price Level Constancy



a 1 percent variation in resources available to the private sector because $g/y = 0.2$ in the stationary state. The government purchase shock requires about a 0.5 percent increase in the money stock, because the neutral private response is to increase output by about 0.5 percent. The price level response is, of course, zero under this policy.

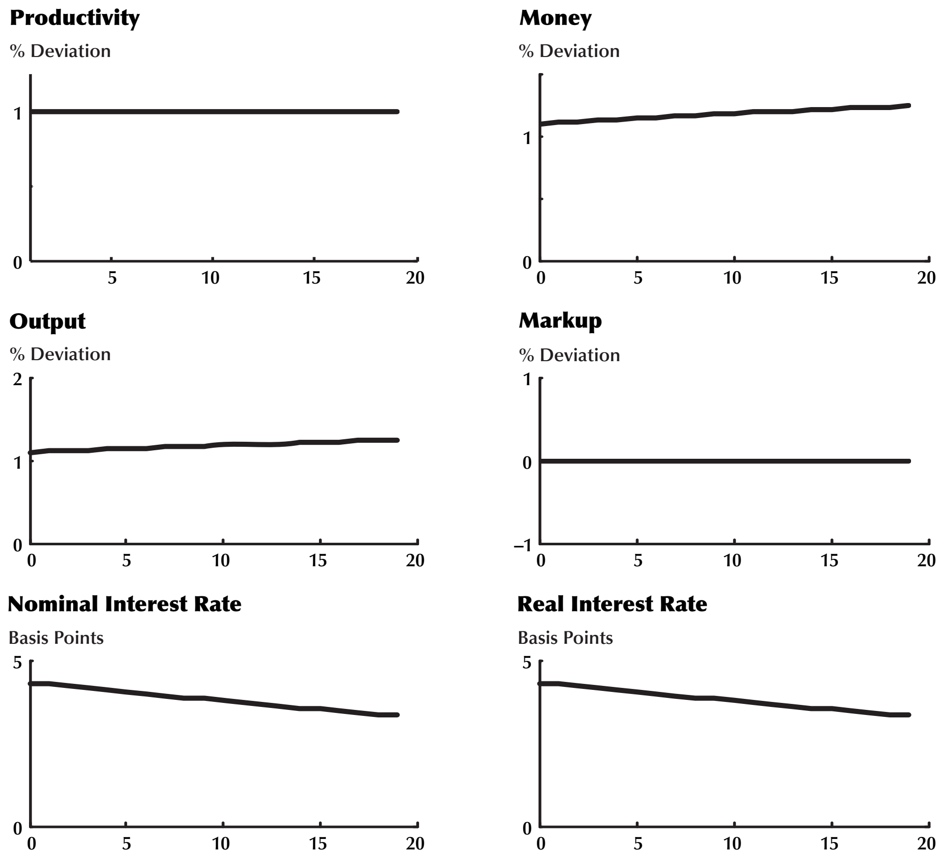
In terms of dynamic responses of model variables, we focus on four variables that will play a role in subsequent discussion. First, output increases because individuals choose to respond in part by working harder and in part by decreasing consumption. Second, the markup charged by firms does not change, simply staying at $\mu = \varepsilon/(\varepsilon - 1)$. Third, the real interest rate rises to stimulate work and to discourage consumption and investment. Fourth, given the absence of expected inflation, the response of the nominal rate is identical to the response of the real rate. The rate increase is quite small, about 20 basis points, despite the presence of substantial investment adjustment costs. This small response reflects the willingness of individuals to substitute consumption and work across time in response to changed intertemporal incentives.

3.2 Response to Productivity

We consider the effects of a permanent 1 percent increase in total factor productivity, as displayed in Figure 2. As we have structured it, this permanent productivity shock has the implication

Figure 2

Effects of a 1 Percent Rise in Total Factor Productivity Under Strict Price Level Targeting



that there is no long-run labor response and there is thus a $1/\alpha = 1.72$ percent effect on the level of output.[7]

The dynamic responses involve a jump in output, followed by a gradual increase toward the new higher steady-state level. The presence of quantitatively important investment adjustment costs means that the equilibrium response of labor to productivity shocks is fairly small. Under flexible prices/strict price level constancy, the average markup does not change in response to this shock, as it also earlier remained constant in the face of government purchase shocks. Variations in the real interest rate in response to this productivity shock are quite small, even in the presence of investment adjustment costs: A large increase in productivity produces only a relatively minor increase in the real interest rate.

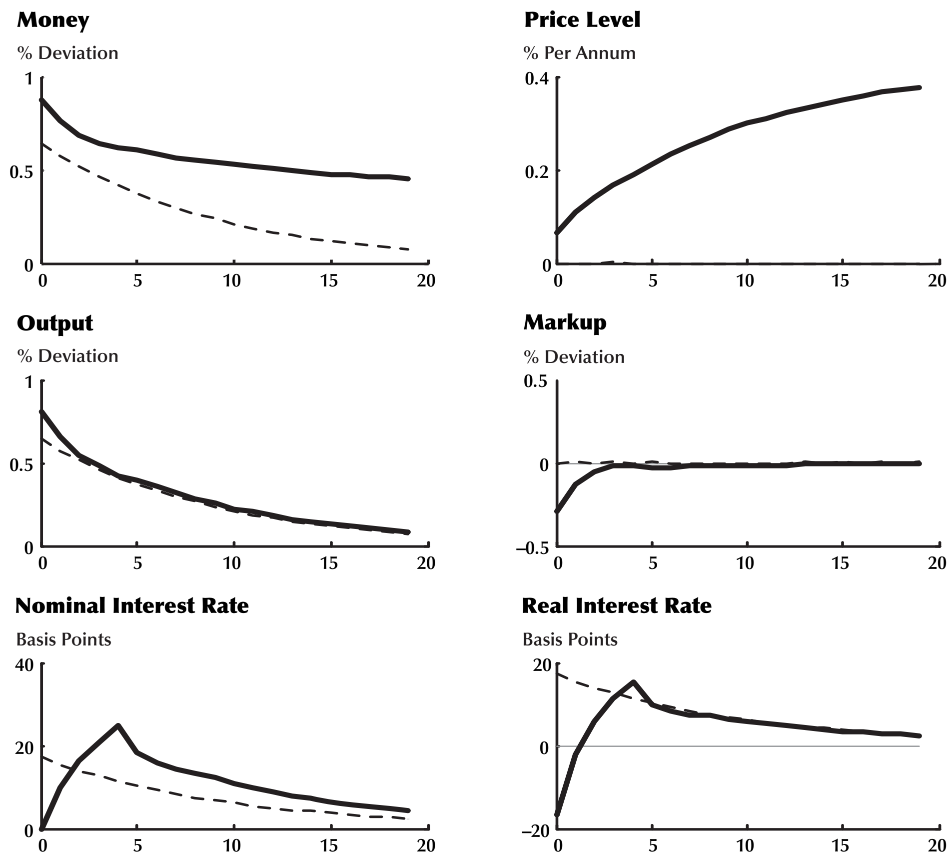
4 RESPONSE UNDER INTEREST RATE RULES

We now consider the dynamic implications of the three standard interest rate rules, studying the response of the macroeconomy to government purchase and productivity shocks in each case. All results are compared with the benchmark results from Figures 1 and 2, which are the results

[7] This zero long-run labor response requires that the permanent productivity shock leaves the g/y ratio unchanged in the long run. To accomplish this, we assume that the government continuously maintains a constant g/y ratio in the face of this shock.

Figure 3

Effects of a Rise in Government Purchases Equal to 1 Percent of GDP Under an Interest Rate Rule that Responds Only to Inflation



that would obtain under strict price level targeting or, equivalently, under flexible prices.

4.1 An Example

As background to analysis of all of the rules, it is useful to consider the following simple equation system. Suppose that the nominal interest rate is given by a Fisher equation,

$$R_t = r_t + E_t\pi_{t+1},$$

as will indeed be the case in our model. Suppose further that the monetary policy rule takes the simple form

$$R_t = r^* + \tau\bar{\pi}_{t-1}$$
$$\bar{\pi}_t = \xi\pi_t + (1-\xi)\bar{\pi}_{t-1}.$$

This policy rule involves no response to deviations of the real rate from its long-run value and no real output term. However, the nominal rate does respond to an exponentially weighted average of inflation.

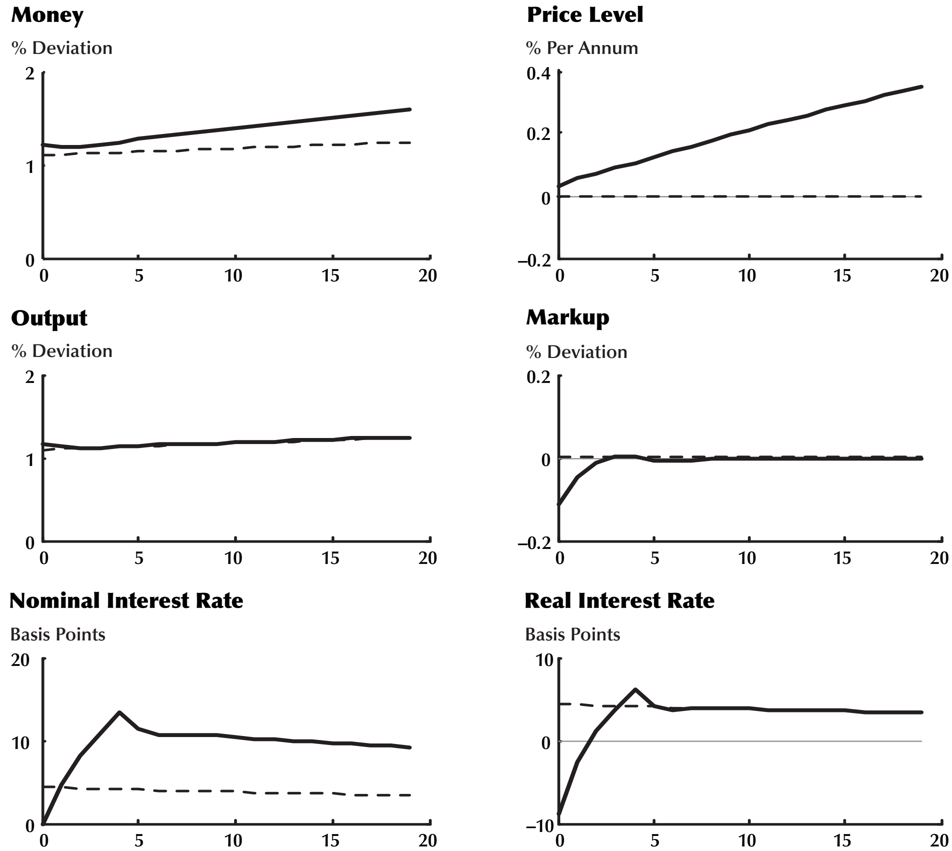
These equations can be solved to yield $E_t\bar{\pi}_{t+1} - (1-\xi)\bar{\pi}_t - \tau\xi\bar{\pi}_{t-1} = -\xi(r_t - r^*)$. With $\tau > 1$, $0 < \xi < 1$ and $\tau < ((2-\xi)/\xi)$, there is a unique stable rational expectations solution,

$$\bar{\pi}_t = \sum_{j=0}^{\infty}\left(\frac{1}{\theta_1}\right)^{j+1}\xi[E_t r_{t+j} - r^*] + \theta_2\bar{\pi}_{t-1},$$

where $\theta_1 > 1$ and $-1 < \theta_2 < 0$. Hence, the "average" inflation rate that enters in the policy rule is positively related to a present value of interest rates and negatively to its own past values.

Figure 4

Effects of a 1 Percent Rise in Total Factor Productivity Under an Interest Rate Rule that Responds Only to Inflation



Next, assume that the real interest rate is exogenous and is governed by a first-order stochastic difference equation, $[r_t - r^*] = \rho[r_t - r^*] + e_t$. Then, it follows that average inflation obeys

$$\bar{\pi}_t = \frac{\xi}{\theta_1 - \rho}[r_t - r^*] + \theta_2 \bar{\pi}_{t-1}$$

and thus that the nominal interest rate evolves according to

$$R_{t+1} = \frac{\tau\xi}{\theta_1 - \rho}[r_t - r^*] + \theta_2 R_t.$$

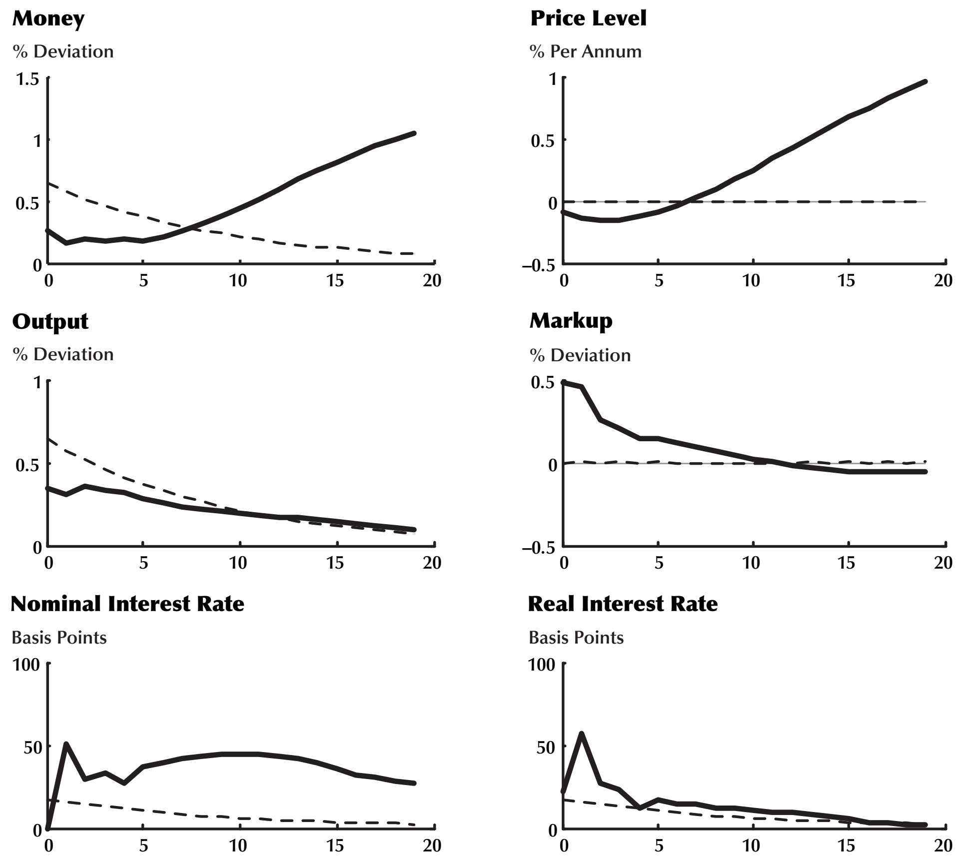
Hence, the nominal interest rate inherits the persistence of the real interest rate, but this is somewhat mitigated due to the presence of the lagged interest rate term that derives from the central bank's concern about average inflation. The solution to this difference equation implies that

$$E_t R_{t+k+1} - E_t R_{t+k} = [\frac{\tau\xi}{\theta_1 - \rho}][\frac{\rho^{k+1} - \theta_2^{k+1}}{\rho - \theta_2}]e_t$$
$$= \frac{\tau\xi}{\tau\xi - \rho[\rho - (1-\xi)]}[\rho^{k+1} - \theta_2^{k+1}]e_t.$$

Accordingly, shocks to the real interest rate exert an amplified influence on the future nominal interest rate under two conditions. First, it must be the case that $\rho > (1 - \xi)$, which is the restriction that the real rate process involves more persistence than the inflation averaging process. Second, it must be the case that one is looking

Figure 5

Effects of a Rise in Government Purchases Equal to 1 Percent of GDP Under the Standard Taylor Rule



sufficiently far ahead that the effects of θ_2 are negligible. Under these two conditions, the real rate then affects the nominal rate with a greater than one-for-one effect because

$$\frac{\tau\xi}{\tau\xi - \rho[\rho - (1-\xi)]} > 1.^{8}$$

This example highlights the fact that interest rate policies can ultimately produce a positive comovement of the real interest rate and expected inflation. That is, one legacy of stabilizing the nominal interest rate in the short run against shifts in the real interest rate is to produce longer-run variations in inflation. Of course, the Fisher equation and the interest rate rules are only part of modern macroeconomic models and it is not the case that the real interest rate is exogenous. However, the mechanisms of this simple model provide insight into aspects of the general equilibrium dynamics discussed hereafter.

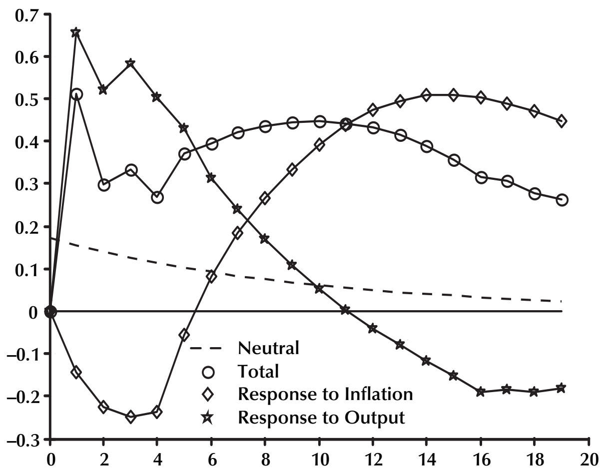
4.2 The "Inflation Only" Rule

We start by considering the implications of an interest rate rule (17), which depends only on recent inflation.

[8] As an example, suppose that $\tau = 1.5$ as in Taylor's analysis and that ξ is set equal to 0.4 as a rough approximation to 4-quarter averaging (the same mean lag). Finally, suppose that the real rate is quite persistent, with $\rho = 0.9$ These parameters imply that $\theta_2 = -0.54$. After 6 quarters (since $0.54^6 = 0.02$, the economy would be left a nominal rate that was about $\tau\xi/[\tau\xi - \rho(\rho - (1-\xi))] = 1.82$ times as high as the real rate.

Figure 6

Decomposition of Nominal Interest Rate Variations in Response to a Government Purchase Shock Under the Taylor Rule



4.2.1 Government Purchases. Figure 3 shows the effects of a rise in government purchases. In line with the suggestions of Brunner, Friedman, and Poole, the dynamic responses in this case involve an exacerbation of the benchmark responses for real output (the neutral responses are the dashed lines in this and subsequent figures). Given that the nominal interest rate is predetermined and the neutral responses (introduced in Figure 1 and repeated as the dashed lines in Figure 3) involve a rise in the real interest rate, there is a tension: There must be either a departure of the real interest rate from its neutral level or a decline in expected inflation. This tension is resolved by a transitory increase in output beyond the benchmark level, which gives individuals an incentive to save and reduces the upward pressure on the real interest rate. As part of the process, the money stock expands more rapidly than in the neutral case, as suggested by Brunner, Friedman, and Poole.

To fit the dynamics together, here and below, we have found it useful to adopt the monetarist perspective, taking the path of the money stock as exogenous and using this path to explain the variations that arise in other variables, including the nominal interest rate (which is governed by the policy rule). To start on this approach, note that Figure 3 implies that the path for money exceeds the neutral policy path. Further, money rises more in the short run than it does in the long run. Given this increase in money, either real income (y) or the price level (P) must increase (or both). With output temporarily high relative to the neutral path, the real interest rate will be lower than the neutral path. But, given that there is expected inflation and the nominal interest rate is predetermined, the real interest rate must actually fall on impact.

After a few quarters or so, real activity is largely on the neutral path. Hence, after this point in the impulse responses, the simple model introduced here previously can be used: The central bank's focus on inflation leads to higher nominal interest rates, which rise more than one-for-one with the real interest rate.

4.2.2 Productivity. Figure 4 shows the dynamic effects of a rise in total factor productivity if the central bank adjusts the interest rate solely in response to recent inflation experience, in line with (17). In this case, the money stock increases relative to the neutral level and then continues to grow for many periods. Similar to the government purchase response, there is a temporary stimulation, although it is smaller and of less-protracted duration. As a consequence, there is a temporary decline in the real interest rate relative to the neutral solution. As was the case with government purchases, output is on the neutral path after about four quarters, so that the simple model can be used to explain the comovement of nominal and real interest rates.

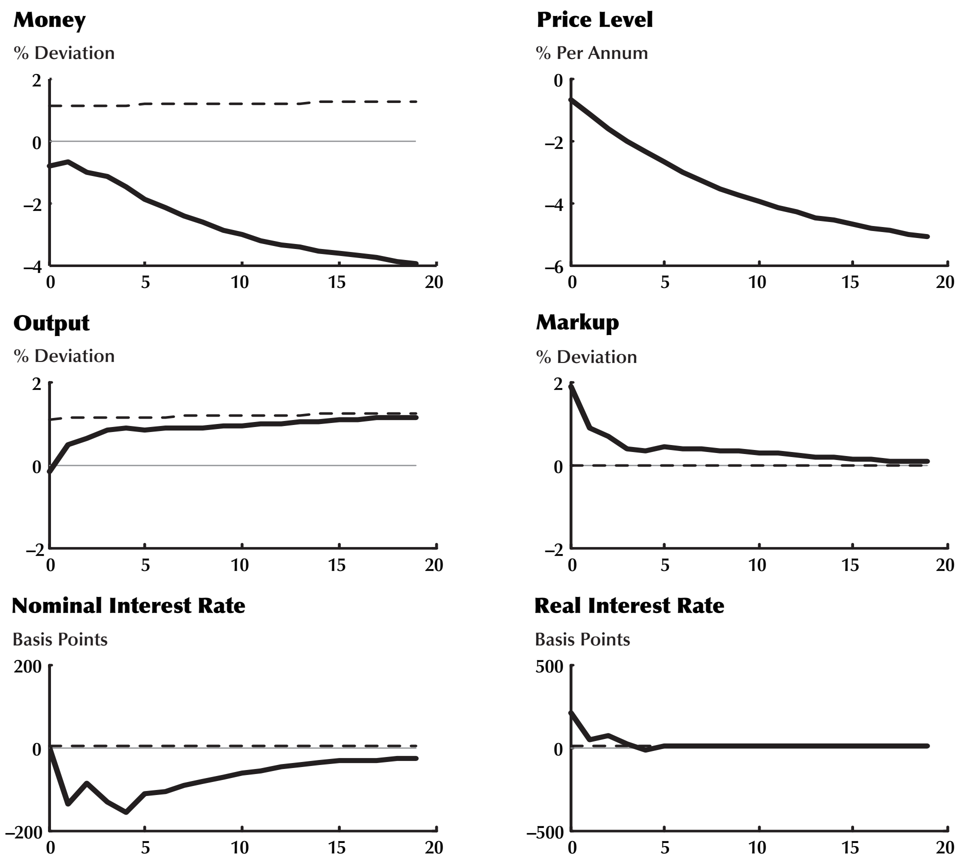
4.2.3 Relationship to Monetarist Critique(s). The two examples studied here show that there can be an "exacerbation" of short-run output responses under an interest rate rule.[9] However, the exacerbation of real responses is small relative to the overall effect of the disturbances on economic activity.

After the initial interval of monetary non-

[9] Quantitatively, the effects are not too large in the current setting, which incorporates no recent mechanisms to increase the amplitude and persistence of monetary shocks.

Figure 7

Effects of a 1 Percent Rise in Total Factor Productivity Under the Standard Taylor Interest Rate Rule



neutrality, there is the comovement of the real rate, the inflation rate, and the nominal interest rate highlighted by the simple model considered in section 4.1. In line with that example, a consequence of the interest rate rule is that the short-run "leaning against" fluctuations in the nominal rate of interest leads to greater medium-term volatility in inflation and the nominal rate, which is related to concerns expressed by Friedman (1968) and Poole (1978).

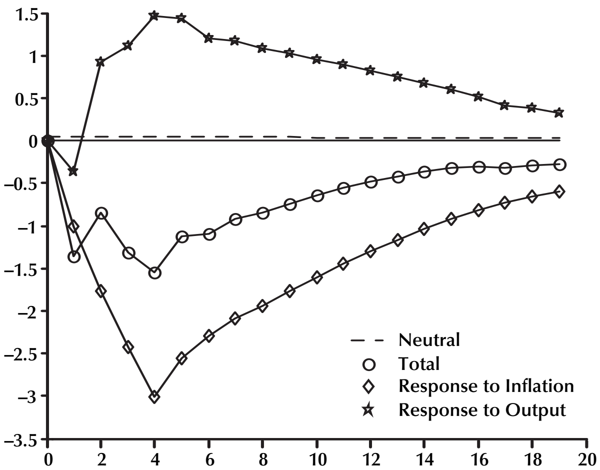
4.3 The Standard Taylor Rule

With a Taylor response to the approximate output gap, ($\tau_2 = 0.5$), a very different pattern of response to government purchase and productivity shocks is displayed in Figures 5 and 7. As indicated by the money supply responses in the first panel of each figure, the central bank does not exacerbate economic fluctuations in the short run under such a Taylor-style interest rate rule. Instead, the interest rate rule works to restrain variations in output, relative to the neutral solution. (Again, this neutral solution is the dashed line in each figure.)

4.3.1 Government Purchases. An increase in government purchases increases output and the real interest rate under the neutral solution, but requires a supporting increase in the money stock. Because there is a reduction in the quantity of money relative to the neutral solution, there is a decline in aggregate demand and output (again

Figure 8

Decomposition of Nominal Interest Rate Variations in Response to a Productivity Shock Under the Taylor Rule



relative to the neutral solution). The decline in demand also induces a subset of firms to adjust prices downward, with a resulting decrease in the price level and an increase in the real markups of the average firm. Because output is temporarily reduced—relative to the neutral benchmark—and because there is a modest expected deflation, the real interest rate rises on impact while the nominal rate is fixed.

There are interesting dynamic elements of the behavior of nominal interest rates under the Taylor rule, displayed in Figure 6. There are two features of importance. First, given the rise in output that takes place due to the increase in government purchases, the Taylor rule coefficient of $\tau_y = 0.5$ implies that there is an initial rise in the nominal interest rate path. Second, after the initial year or so, high inflation is an inheritance of a shock to government purchases, as in our example noted previously.

4.3.2 Productivity. A permanent increase in productivity raises output permanently and the real interest rate temporarily under the neutral solution, but again requires a supporting increase in the money stock. The Taylor rule dramatically alters the effects of this shock relative to the neutral solution: It leads to a sharp rise in the real interest rate as output adjusts only gradually toward its new higher level. In our context, then, the Taylor rule is consistent with the effects of productivity shocks that Galí (1999) displays in a sticky price model with an exogenous money stock.

By looking at the path of the quantity of money, the short-run parts of the dynamic responses are again relatively easy to understand. When the productivity shock takes place, the quantity of money immediately declines. This leads to a decline in the level of output even though productivity has risen, a decline in the price level, and a sharp rise in the average markup charged by firms. Since output is temporarily low relative to the neutral path (and relative to the long-run level), there is a substantial increase in the real interest rate. The Taylor rule is consistent with these changes because the nominal interest rate is held fixed, while expected deflation occurs.

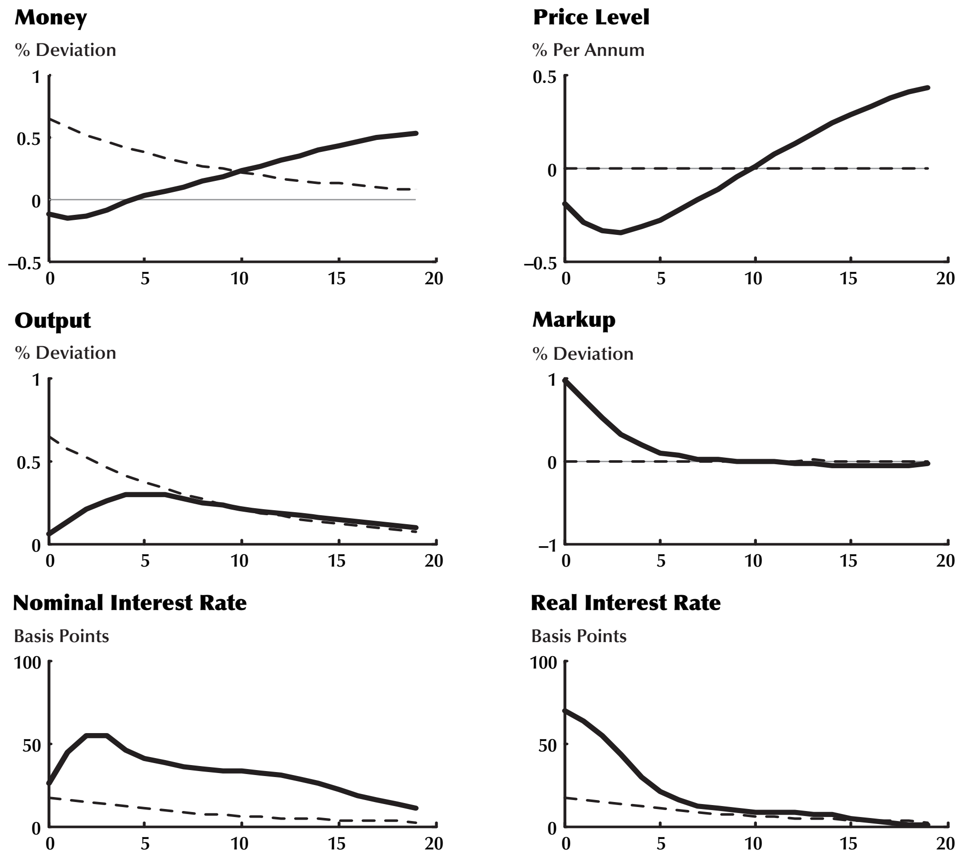
The dynamic response to inflation and real activity built into the Taylor rule also means that there are longer-lasting effects of the productivity shock on the path of inflation and the nominal interest rate. The productivity shock gives rise to a sustained period of deflation, during which the nominal interest rate is low. Figure 8 shows that the dominant quantitative effect on interest rate policy is from inflation, but that there is also a contribution from output being high relative to the central bank's response to a slowly evolving measure of the output gap.

4.3.3 Relationship to Monetarist Critique(s). These responses to government purchase and productivity shocks under the Taylor rule illustrate that interest rate rules can "stabilize" rather than "exacerbate" responses to real shocks. However, the stabilization is likely undesirable, because there is a presumption in these models that the natural rate solution is approximately optimal (see Goodfriend and King, 1997, and Woodford, 2003).

These two examples also show that government purchase and productivity shocks, which both raise the real interest rate, can exert different effects on inflation. Government purchases first

Figure 9

Effects of a Rise in Government Purchases Equal to 1 Percent of GDP Under the Orphanides-Wieland Rule



decrease and then increase the inflation rate; productivity movements decrease it under the Taylor rule. Because these variations are largely anticipated after a year or so, they are manifest in high or low nominal interest rates during the latter portions of the dynamic responses.

4.4 An Estimated Rule

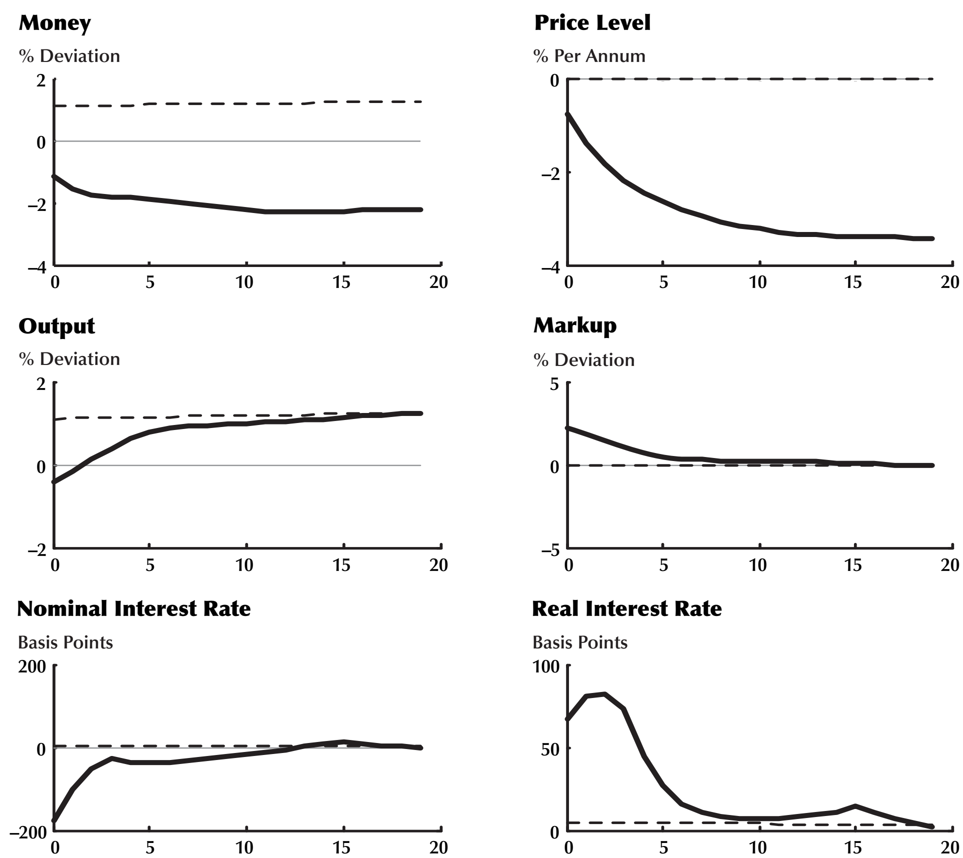
There has been much recent interest in estimating interest rate rules for the United States and other countries. Typically, these studies turn up support for one or more interest rate lags and one or more output gap lags. The estimated rule of Orphanides and Wieland (1998) is representative in this regard. Figures 9 and 10 provide an indication of the effects of such an estimated rule within our model.

In terms of government purchases, the responses in Figure 9 look much like "a smoother Taylor model," with the main features preserved from that simpler specification.[10] The Orphanides-Wieland rule leads output to respond less to government purchases in the short run than it would under the neutral specification. When there are increases in demand, the Orphanides-Wieland rule also leads to a short period of declining prices followed by inflation.

[10] One exception is that the nominal interest rate is not predetermined under the Orphanides-Wieland rule because the central bank responds contemporaneously to output.

Figure 10

Effects of a 1 Percent Rise in Productivity Under the Orphanides-Wieland Rule



In terms of productivity shocks, the responses in Figure 10 also inherit the main features of the Taylor rule, although yielding a smoother and more-protracted pattern of responses. The Orphanides-Wieland interest rate rule retards the expansion of output from the productivity shock, inducing a rise in the markup and the real interest rate in the early stages of the expansion. A decline in the money stock and the price level accompany the productivity expansion.

5 CONCLUSIONS

We view the analysis in this article as our first step in reassessing the continuing relevance of the monetarist critique of interest rate rules to two important topics. First, we are interested in whether these concerns are important for understanding aspects of the macroeconomic history of the United States and other countries. That is, did the use of specific interest rate rules for monetary policy exacerbate or stabilize the business cycle fluctuations that would have arisen from real shocks? Second, we are interested in whether these concerns are important to the design of good operating rules for monetary policy. That is, how large are the departures of specific interest rate rules from optimal policies?

Our analysis highlights the role that real factors can play in both of these investigations. With sticky prices, interest rate rules can produce real outcomes that diverge from the real business cycle responses that would be delivered by neutral policy, with the policy rules either exacerbating or stabilizing the real cycle. As suggested by monetarist economists—like Brunner, Friedman, and Poole—variations in the money stock are one set of indicators of these departures. However, our modern model specifies that the relevant monetary indicator is the gap between the money supply response under a specific interest rate rule and that which would occur under a neutral policy that effectively accommodated real shocks. In addition, our modern model suggests that interest rate rules increase the lower-frequency variability of inflation.

REFERENCES

Baxter, Marianne and King, Robert G. "Fiscal Policy in General Equilibrium." *American Economic Review*, June 1993, *83*(3), pp. 315-34.

Brunner, Karl. "Commentary on Monetary Economics: An Interview with Karl Brunner." *The Banker*, July 1978, *128*(629), pp. 20-23. Reprinted in Federal Reserve Bank of St. Louis *Review*, November 1978, *60*(11), pp. 2-7.

Calvo, Guillermo A. "Staggered Prices in a Utility-Maximizing Framework." *Journal of Monetary Economics*, 1983, *12*(3), pp. 383-98.

Dixit, Avinash K. and Stiglitz, Joseph E. "Monopolistic Competition and Optimum Product Diversity." *American Economic Review*, 1977, *67*(3), pp. 297-308.

Dotsey, Michael. "Structure from Shocks." Federal Reserve Bank of Richmond *Economic Quarterly*, Fall 2002, *88*(4), pp. 1-13.

Friedman, Milton. "The Role of Monetary Policy." *American Economic Review*, May 1968, *58*, pp. 1-17.

Friedman, Milton. "The Optimum Quantity of Money," in *The Optimum Quantity of Money and Other Essays*. Chicago: Aldine, 1969.

Friedman, Milton. "Monetarism in Rhetoric and Practice," in Albert Ando, ed., *Monetary Policy in Our Times*. Proceedings of the First International Conference held by the Institute for Monetary and Economic Studies of the Bank of Japan. Cambridge, MA: MIT Press, 1985, p. 21.

Galí, Jordi. "Technology, Employment, and the Business Cycle: Do Technology Shocks Explain Aggregate Fluctuations?" *American Economic Review*, March 1999, *89*(1), pp. 249-71.

Goodfriend, Marvin and King, Robert G. "The New Neoclassical Synthesis and the Role of Monetary Policy," in Ben S. Bernanke and Julio J. Rotemberg, eds., *NBER Macroeconomics Annual 1997*. Cambridge, MA: MIT Press, 1997, pp. 231-83.

Hayashi, Fumio. "Tobin's Marginal q and Average q: A Neoclassical Interpretation." *Econometrica*, January 1982, *50*(1), pp. 213-24.

Kimball, Miles S. "The Quantitative Analytics of the Basic Neomonetarist Model." *Journal of Money, Credit, and Banking*, 1995, *27*(4 Part 2), pp. 1241-77.

King, Robert G. and Watson, Mark W. "Money, Prices Interest Rates and the Business Cycle." *Review of Economics and Statistics*, February 1996, *78*(1), pp. 35-53.

King, Robert G. and Wolman, Alexander L. "Inflation Targeting in a St. Louis Model of the 21st Century." Federal Reserve Bank of St. Louis *Review*, May/June 1996, *78*(3), pp. 83-107.

Lucas, Robert E. Jr. "Econometric Policy Evaluation: A Critique." *Carnegie-Rochester Conference Series on Public Policy*, 1976, (1), pp. 19-46.

Orphanides, Athanasios and Wieland, Volker. "Price Stability and Monetary Policy Effectiveness when Nominal Interest Rates Are Bounded at Zero." Finance and Economics Discussion Series, 1998-35. Board of Governors of the Federal Reserve System, 1998; www.federalreserve.gov/pubs/feds/1998/199835/199835abs.html.

Poole, William. *Money and the Economy: A Monetarist View*. Reading, MA: Addison-Wesley, 1978.

Rotemberg, Julio J. "The New Keynesian Microfoundations," in Stanley Fischer, ed., *NBER Macroeconomics Annual 1987*. Cambridge, MA: MIT Press, 1987, pp. 63-129.

Taylor, John B. "Discretion Versus Policy Rules in Practice." *Carnegie-Rochester Conference Series on Public Policy*, 1993, *39*, pp. 195-214.

Woodford, Michael. *Interest and Prices: Foundations of a Theory of Monetary Policy*. Princeton, NJ: Princeton University Press, 2003.

Yun, Tack. "Nominal Price Rigidity, Money Supply Endogeneity, and Business Cycles." *Journal of Monetary Economics*, April 1996, *37*, pp. 345-70.

Commentary

Julio J. Rotemberg

The King and Lin (2005) paper analyzes the desirability of various interest rate rules in a standard model of sticky prices. A simplified version of this standard model consists of (PS), a price-setting rule, and (IS), an investment-equals-savings curve. These can be written as

$$\text{(PS)} \qquad \pi_t = \kappa\left(y_y - y_t^e\right) + \beta E_t \pi_{t+1} + \omega_t$$

$$\text{(IS)} \quad y_y - y_t^e = E_t\left(y_{t+1} - y_{t+1}^e\right) - \sigma\left(i_t - E_t \pi_{t+1} - v_t\right),$$

where π_t, i_t, y_t represent the observable inflation rate, nominal interest rate, and log of output at t, respectively; y_t^e represents the efficient level of output; and ω_t and v_t represent stochastic disturbances.

The paper considers a monetary authority that sets interest rates as a function of the economy's history and evaluates how these rules perform when the economy is subject to two types of shocks. The first of these shocks is a technology shock. This shock affects a_t in the production function

$$(1) \qquad y_t = \log\left(a_t f\left(k_t, n_t\right)\right),$$

where n_t and k_t represent the labor input and capital, respectively. Such a shock obviously affects the efficient level of output, y_t^e. In general, it would also affect v_t—that is, the real interest rate that is consistent with ensuring that the efficient level of output, y_t^e, is demanded at t, given that the efficient level of output will be demanded in the future.

The second shock considered by King and Lin is a shock to government purchases, and this shock too affects y_t^e and v_t. Because neither of these shocks affects ω_t in the (PS) equation, it is apparent from the inspection of (PS) and (IS) that these shocks are consistent with keeping inflation constant and output, y_t, equal to the efficient level, y_t^e. What is needed for such a policy is simply that the nominal rate of interest be set equal to a constant inflation rate, $\bar{\pi}$, plus v_t. Such a policy, particularly when coupled with a low value for $\bar{\pi}$, appears desirable from several different points of view (not the least that it avoids many of the distortions introduced by price rigidity). I thus agree with King and Lin that it provides a natural benchmark against which other interest rate rules can be judged.

While it is a good benchmark, it seems difficult to implement such a rule in practice because the Wicksellian natural rate of interest, v_t, is generally not observable right away. This difficulty provides a rationale for rules where the interest rate responds to the economy's history. Two of the rules considered in this paper are based on the equation

$$(2) \qquad i_t = \bar{i}, + \tau_1 \frac{1}{4} \sum_{j=1}^{4} \pi_{t-j} + \tau_2\left(y_{t-1} - y_t^*\right),$$

where y_t^* is a measure of trend output. With τ_1 = 1.5 and τ_2 = 0.5, this is the rule that Taylor (1993) shows to be similar to actual U.S. policy. King and Lin also consider a variant where τ_1 remains equal to 1.5 and τ_2 is set to zero. Because the central bank responds only to inflation and not to out-

Julio J. Rotemberg is a professor at the Harvard Business School.

Federal Reserve Bank of St. Louis *Review*, July/August 2005, *87*(4), pp. 531-35.

put in this latter rule, King and Lin call it an "inflation only" rule.

The paper spends some time discussing monetarist concerns. Monetarists criticized policies that kept interest rates stable on the ground that they would lead government purchase shocks (or, rather, IS shocks in general) to have too large an effect on output. Policies that stabilized money growth, they argued, would lead interest rates to increase in response to such shocks and thereby stabilize output.

Inspired by these concerns, King and Lin ask themselves whether the interest rate rules that they consider lead to excessive output stability or not. It is important to stress, though, that the monetarists critique was applied to attempts by the monetary authority to stabilize interest rates, whereas interest rates vary over time in the rules studied by King and Lin. An even more important difference between the framework underlying the monetarist analysis and that of King and Lin is that the former saw the stability of output as desirable, whereas the latter focus on a model where even changes in government purchases ought to lead to changes in output.

These differences are sufficiently large that the issue of whether a particular interest rate rule "destabilizes output" relative to the benchmark turns out to be quite different from the issue of whether this particular interest rate rule is desirable in terms of leading to an outcome that is close to the benchmark. The inflation-only rule, in particular, "destabilizes output" in the sense that the failure of interest rates to rise immediately when there is a positive shock to government purchases leads output to rise by more than it would under the benchmark rule. By contrast, the Taylor (1993) rule "stabilizes output" in the sense that its tendency to let interest rates rise when output is above its trend leads output to respond less to both government purchases and to technology shocks than it does in the benchmark for both.

At the same time, the inflation-only rule leads to paths of output that are quite close to those of the benchmark, particularly in response to technology shocks. Thus, this rule appears more desirable than rules that let interest rates respond to output as well, even though there is a sense in which the inflation-only rule "destabilizes" output. This is not altogether surprising. King and Lin focus on a setup where output ought to vary (with y_t^e), while inflation ought to be stable, and it thus makes sense that a rule that focuses only on inflation stabilization is superior to one that tries to stabilize output as well.

The question this raises, however, is whether the output fluctuations one would actually observe in the economy if inflation were stabilized would correspond to fluctuations in the efficient level of output, y_t^e. The combination of equations (PS) and (IS) is silent on this issue because of the existence of ω_t. Fluctuations in ω_t lead to fluctuations in inflation even in the case where output is equal to y_t^e so that, naturally, they would imply undesirable output fluctuations if monetary policy were able to stabilize inflation. Thus, as emphasized by Giannoni and Woodford (2003), fluctuations in ω_t justify efforts by the central bank to stabilize output.

This renders it important to ascertain the extent to which the economy is subject to the shocks to technology and government purchases studied by King and Lin and the extent to which the economy is subject instead to shocks to ω_t. Equations (IS) and (PS) cannot, by themselves, be used to make this judgment because they do not separately identify the roles of ν_t, y_t^e, and ω_t. However, it would seem possible to use information on actual government purchases to identify fiscal shocks and information from an aggregate production function, such as (1), to identify technology shocks.

There is indeed an extensive literature seeking to measure the effect of fiscal shocks on aggregate activity (see Blanchard and Perotti, 2002, for a relatively recent example). One issue in this literature is that some components of government spending, and particularly those related to the welfare state, respond to fluctuations in gross domestic product (GDP) that are due to other causes. However, changes in military purchases in the United States after World War II are plausibly free from this problem, and Rotemberg and Woodford (1992) use these changes to construct a measure of government purchase shocks. They show, moreover, that the changes in GDP that follow these shocks are broadly consistent with those of a competitive model where markups are constant. On the other hand, they also show that real wages increase after an increase in military purchases, and this suggests that increases in military purchases lower markups.

By contrast, King and Lin's simulations show that when the monetary authority uses a Taylor rule, markups strongly rise. The reason is not hard to understand. In the standard neoclassical model as well as the King and Lin model, the principal reason output rises in response to an increase in government purchases is that this increase makes people feel poorer and increases their labor supply (reduces their consumption of leisure). There is thus downward pressure on real wages and on prices. The downward pressure on prices tends to translate into increased markups when prices are sticky, particularly if the monetary authority raises interest rates in response to increases in output.

One reason to be skeptical of this mechanism is that it requires that workers/consumers respond to shocks to their future wealth, and there is considerable evidence that people often do not adjust their consumption until their flow of income changes. Insofar as shocks that increase government purchases do not lead to immediate increases in taxes, the reduction in the flow of income is deferred and this might well also defer the increase in labor supply.

This argument as well as the evidence of Rotemberg and Woodford (1992) concerning the response of real wages both suggest that the increase in output that follows government purchase shocks may well be due to reductions in desired markups. This would mean that such shocks affect ω and might thus justify a monetary policy that is geared toward output stabilization.

The identification of technology shocks presents substantial difficulties. An approach that is known to be problematic is to equate technological change with computed Solow residuals. The problem is that short-term changes in the Solow residuals and movements in output induced by nontechnological factors can be connected through several mechanisms, including variations in labor effort and the existence of increasing returns to scale. A potentially more-promising method for detecting changes in the technology factor, a_t, is to use information about changes in the long-run level of output or labor productivity. Galí (1999) and Galí and Rabanal (2005), in particular, use a bivariate vector autoregression (VAR) that includes changes in hours and labor productivity and identifies technology shocks as those shocks that are orthogonal to the shocks that have no long-run effect on the level of productivity. Given the specification, these shocks clearly *do* affect productivity in the long run. The technology shocks identified by this method are initially followed by declines in hours. Because hours and output move together over the business cycle, it appears that this method does not give an important role for technology shocks in business fluctuations.

A source of concern with this method of identifying technology shocks is that a bivariate VAR in hours and productivity growth has difficulty explaining long-term changes in productivity growth. In the U.S. nonfarm business sector, the compound annual growth rate of productivity was 2.75 percent from 1947:Q1 to 1973:Q4, whereas it was 1.66 percent from 1974:Q1 to 1992:Q4. To see whether Galí and Rabanal's (2005) statistical model is consistent with this, I first estimated a VAR explaining changes in nonfarm hours and nonfarm productivity from 1949:Q2 to 2002:Q4 using four lags of the two variables. Using these estimated parameters, I generated 10,000 histories of the two endogenous variables by randomly generating residuals that had the same variance covariance matrix as those of the estimated VAR. For each history, I computed the compound rate of growth of labor productivity from observation 1 to observation 107 as well as the compound rate of growth from observation 108 to observation 183. The maximum difference between these two compound rates of growth across these 10,000 histories was 0.4 percent, which is considerably below the 1.1 percent difference in compound growth rates observed in the data between the first and second periods described here previously.

This difficulty with a bivariate VAR in first differences echoes some of the well-known difficulties posed by the post-1973 productivity slowdown. One way of capturing this slowdown is to suppose that there was a break in the economy's mean growth rate at some point in the early 1970s. This would imply that the population of the United States suffered a huge wealth loss at that time. Given the lack of a precipitous reduction in asset prices at the time, the idea that the economy's underlying rate of growth fell suddenly is not wholly convincing. An alternative view is to suppose that technological progress is much smoother.

This is the approach taken in Rotemberg (1999, 2003), where I fit a smooth trend, d_t, to output by minimizing

$$\sum_{t=2}^{T-1}\left[d_t - 2d_{t-1} + d_{t-2}\right]^2 + (1/\lambda)\sum_{t=1+k}^{T}\left(y_t - d_t\right)\left(y_{t-k} - d_{t-k}\right), \tag{3}$$

where the parameter y_t is the output level at t and λ is set at the lowest possible value that ensures that

$$\sum_{t=k+v}^{T-k-v}\left(y_t - d_t\right)\left[d_{t+v} - 2d_t + d_{t-v}\right] = 0. \tag{4}$$

The resulting trend is quite smooth and has two additional properties. The first is that the covariance of the cycle (the difference between y_t and d_t) at t and $t-k$ is small. For $k = 16$, this ensures that cycles are essentially over after four years. The second property is that the cycle at t is orthogonal to the deviation of the trend at t from the average of its values at $t+v$ and $t-v$. This ensures that the trend does not behave in a way that is similar to that of the cycle itself.

Rotemberg (2003) shows that a process for technical progress that is consistent with microeconomic evidence on diffusion lags generates movements in GDP that are quite similar to those of the trend that is constructed by this method. This means that, if technical progress does indeed take this form, stabilizing output relative to trend becomes a valid goal for monetary policy once again. On the other hand, the computation of trend output, y_{t^*}, may well be complicated in this case. In particular, Orphanides and van Norden (2002) argue that $y_t - y_{t^*}$ is particularly difficult to compute in real time when y_{t^*} is smooth and yet different from a linear trend.[1]

This raises the question whether, in practice, the Federal Reserve has had difficulty dealing with recent episodes of changes in technical progress. While a thorough answer to this question is well beyond the scope of this comment, it is worth studying briefly the increase in productivity that took place in the United States in the 1990s. The idea is to see whether allegiance to a monetary rule of the type used by Taylor (1993) to describe actual Federal Reserve practice caused difficulties when technical progress accelerated in the 1990s. To see that it could, note that (2) leads to a systematic relation between inflation, the real rate of interest, and the departure of output from y^*. In particular, treating inflation in adjacent periods as being essentially equal to π_t, it yields

$$\pi_t = 2\left(i_t - \pi_t - i^*\right) - \left(y_t - y_t^*\right). \tag{5}$$

Over long periods of time, one would not expect the central bank to have much influence on the real interest rate, which appears on the right-hand side of this equation. Moreover, in line with the King and Lin model, one would expect a rise in productivity growth induced by technical progress to raise this real rate. In practice, it seems that trend real rates fell over the 1990s, however. If one applies the detrending method of Rotemberg (1999) to the real interest rate based on Treasury bill rates and the consumer price index (CPI) from 1981:Q1 to 2003:Q4, one finds that the trend real rate fell from 2.52 percent per annum in 1991:Q1 to 1.35 percent per annum in 1999:Q2. To compute the change in $y_t - y_{t^*}$, I use Taylor's (1993) method for obtaining y_{t^*}. This involves the fitting of a linear trend for the logarithm of real GDP from 1984:Q1 to 1993:Q2. The result is that $y_t - y_{t^*}$ went from being –1.8 percent in 1991:Q1 to 3.9 percent in 1999:Q2. Using (5), this increase in "cyclical" output over nine years, together with the corresponding reduction in the trend real rate of interest, ought to have lowered the inflation rate by about 8 percent per year. In practice, the 12-month CPI inflation rate was essentially the same in 1991:Q1 as it was in 1999:Q2.

---

[1] For a proposal for dealing with this problem, see Orphanides and Williams (2002).

The precise numbers that belong in the above calculation are obviously subject to a great deal of uncertainty. It remains the case, however, that the "natural real rate of interest," v_t, appears to have declined in the 1990s, whereas "cyclical output" would have increased a great deal if a naive linear trend had been used to compute the "natural" level of output. The use of a mechanical rule setting interest rates as a function of inflation and a naive measure of cyclical output would thus have tended to reduce inflation considerably. That inflation remained stable suggests that the Federal Reserve was able to deal effectively with these underlying changes in the economy. One reason for this may have been that, in spite of the difficulties noted by Orphanides and van Norden (2002) for the computation of trend output in real time, the Federal Reserve was able to obtain an accurate reading of this particular trend change. Insofar as the central bank can be counted on to track changes in trend output accurately, and insofar as technical progress affects only such trends, efforts at stabilizing cyclical output are more likely to be beneficial.

REFERENCES

Blanchard, Olivier and Perotti, Roberto. "An Empirical Characterization of the Dynamic Effects of Changes in Government Spending and Taxes on Output." *Quarterly Journal of Economics*, November 2002, *117*(4), pp. 1329-68.

Galí, Jordi. "Technology, Employment, and the Business Cycle: Do Technology Shocks Explain Aggregate Fluctuations?" *American Economic Review*, March 1999, *89*(1), pp. 249-71.

Galí, Jordi and Rabanal, Pau. "Technology Shocks and Aggregate Fluctuations: How Well Does the RBC Model Fit Postwar U.S. Data?" in Mark Gertler and Kenneth Rogoff, eds., *NBER Macroeconomics Annual 2004*. Cambridge, MA: MIT Press, 2005, pp. 225-88.

Giannoni, Marc P. and Woodford, Michael. "Optimal Interest-Rate Rules: II. Applications." NBER Working Paper No. 9420, National Bureau of Economic Research, 2003.

King, Robert G. and Lin, Mau-Ting. "Reassessing the Monetarist Case Against Interest Rate Rules." Federal Reserve Bank of St. Louis *Review*, July/August 2005, *87*(4), pp. 513-30.

Orphanides, Athanasios and Williams, John C. "Robust Monetary Policy Rules with Unknown Natural Rates." *Brookings Papers on Economic Activity*, 2002, (2), pp. 63-118.

Orphanides, Athanasios and van Norden, Simon. "The Unreliability of Output Gap Estimates in Real Time." *Review of Economics and Statistics*, November 2002, *84*(4), pp. 569-83.

Rotemberg, Julio J. "A Heuristic Method for Extracting Smooth Trends from Economic Time Series." NBER Working Paper No. 7439, National Bureau of Economic Research, December 1999.

Rotemberg, Julio J. "Stochastic Technical Progress, Smooth Trends, and Nearly Distinct Business Cycles." *American Economic Review*, December 2003, *95*(5), pp. 1543-59.

Rotemberg, Julio J. and Woodford, Michael. "Oligopolistic Pricing and the Effects of Aggregate Demand on Economic Activity." *Journal of Political Economy*, December 1992, *100*(6), pp. 1153-207.

Taylor, John B. "Discretion versus Policy Rules in Practice." *Carnegie-Rochester Conference Series on Public Policy*, December 1993, *39*, pp. 195-214.

Productivity and the Post-1990 U.S. Economy

Ellen R. McGrattan and Edward C. Prescott

In this paper, the authors show that ignoring corporate intangible investments gives a distorted picture of the post-1990 U.S. economy. In particular, ignoring intangible investments in the late 1990s leads one to conclude that productivity growth was modest, corporate profits were low, and corporate investment was at moderate levels. In fact, the late 1990s was a boom period for productivity growth, corporate profits, and corporate investment.

Federal Reserve Bank of St. Louis *Review*, July/August 2005, *87*(4), pp. 537-49.

The standard measure of productivity is gross domestic product (GDP) per hour worked. The thesis of this paper is that this measure of productivity is not a good measure of actual output produced per hour worked, which we call *economic productivity*. The reason is that output is understated by GDP because many investments are not accounted for in the U.S. Department of Commerce's Bureau of Economic Analysis (BEA) measure of product. If the importance of these unaccounted investments relative to GDP remained constant over time, growth in GDP per hour would be equal to growth in economic productivity. But in the post-1990 U.S. economy, the relative importance of these investments varied a lot. We find that excluding them in the measure of U.S. output leads to a large underestimate of productivity growth in the late 1990s.

In this paper, these unaccounted investments will be called *intangible investments*.[1] They are expenditures that increase future profits but, by national accounting rules, are treated as an operating expense rather than as a capital expenditure. Examples include advertising, research and development, and, most important of all, investments in building organizations. Most intangible investments are not directly observable, but they can be inferred using standard growth theory and data from the U.S. national income and product accounts (NIPA). We do this and show that movements in accounting and economic measures of productivity are very different during the 1990s. In particular, we find that productivity growth prior to 1997 was even weaker than suggested by GDP per hour worked, that there was a productivity boom in the late 1990s, and that productivity growth returned to its low level subsequent to the boom.

Our accounting has other implications, in particular, for corporate profits and corporate investment. In the late 1990s output boom, the corporate profit share reported by the BEA was low. (See U.S. Department of Commerce, 1929-2004.) A low profit share is not typical in booms.

[1] We use this term because the bulk of the investments are not tangible.

Ellen R. McGrattan is a monetary advisor at the Federal Reserve Bank of Minneapolis and an adjunct professor of economics at the University of Minnesota. Edward C. Prescott is the W.P. Carey Professor of Economics at Arizona State University and a senior monetary advisor at the Federal Reserve Bank of Minneapolis. Data are also available at the website of the Federal Reserve Bank of Minneapolis: www.minneapolisfed.org. The authors thank Ricardo Caballero for helpful comments on an earlier draft. They also thank the National Science Foundation for financial support. The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.

The reason the corporate profit share fell in the boom is simple accounting: Accounting profits understated economic profits because corporations were making large intangible investments in the late 1990s that they expensed. Adding intangible investments to accounting profits and to accounting investment implies a very different picture of the U.S. economy than is evident in the BEA data because adjusted corporate profit and corporate investment shares were both high.

THE MODEL ECONOMY

In this section, we present a version of the model economy that we used to analyze secular movements in corporate equity values. (See McGrattan and Prescott, 2004.) Here, we use the model to compare accounting and economic measures of key aggregate statistics.

The economy is populated by a large number of identical, infinitely lived households. They make decisions about consumption, labor supply, and saving. These decisions are event contingent, where the events are generated by a finite-state Markov chain with stationary transition probabilities. The period t state is $s_t \in S$.

Preferences of the stand-in household are ordered by

$$E\sum_{t=0}^{\infty}\beta^t U(C_t/N_t, L_t/N_t), \tag{1}$$

where C_t is total consumption, L_t is total labor supply, and N_t is the working-age population.

There are two stand-in firms, one for the corporate sector and one for the noncorporate sector. The constant-returns technology for the corporate sector, sector 1, is given by

$$Y_{1,t} = f^c(K_{1m,t}, K_{1u,t}, L_{1,t}, s_t)$$

$$K_{1m,t+1} = (1-\delta_{1m})K_{1m,t} + X_{1m,t}$$

$$K_{1u,t+1} = (1-\delta_{1u})K_{1u,t} + X_{1u,t}$$

$$\hat{K}_{1m,t+1} = [(1-\hat{\delta}_{1m})\hat{K}_{1m,t} + (1-\hat{\delta}_{1x})X_{1m,t}]/(1+\pi_{t,t+1}),$$

where $Y_{1,t}$ is the output of the sector, $K_{1m,t}$ is the beginning-of-period stock of measured capital, $K_{1u,t}$ is the beginning-of-period stock of unmeasured capital, $\hat{K}_{1m,t}$ is the beginning-of-period book value of the stock of measured capital, $L_{1,t}$ is the labor input, $X_{1m,t}$ is new investment in measured capital, $X_{1u,t}$ is new investment in unmeasured capital, and $\pi_{t,t+1}$ is the inflation rate between t and $t+1$. Later, we use the fact that f^c has a unit elasticity between capital and labor, with the capital share equal to θ.

Stocks $K_{1m,t}$ and $\hat{K}_{1m,t}$ can be different if tax rules allow for differences between depreciation for taxes and actual economic depreciation. They can also be different if there is inflation. The stocks of measured and unmeasured capital depreciate at rates δ_{1m} and δ_{1u}, respectively. For tax purposes, capital consumption allowances are equal to $\hat{\delta}_{1m}\hat{K}_{1m,t} + \hat{\delta}_{1x}X_{1m,t}$ and can exceed $\delta_{1m}K_{1m,t}$ because of accelerated depreciation allowances or allowances by the Internal Revenue Service for expensing tangible investments.

The constant-returns technology for the noncorporate sector, sector 2, is

$$Y_{2,t} = f^{nc}(K_{2m,t}, L_{2,t}, s_t)$$

$$K_{2m,t+1} = (1-\delta_{2m})K_{2m,t} + X_{2m,t}$$

$$\hat{K}_{2m,t+1} = [(1-\hat{\delta}_{1m})\hat{K}_{2m,t} + (1-\hat{\delta}_{1x})X_{2m,t}]/(1+\pi_{t,t+1}),$$

where $Y_{2,t}$ is the output of the sector, $K_{2m,t}$ is the beginning-of-period stock of measured capital, $\hat{K}_{2m,t}$ is the beginning-of-period book value of the stock of measured capital, $L_{2,t}$ is the labor input, and $X_{2m,t}$ is new investment in measured capital. Later, we use the fact that f^{nc} has a unit elasticity between capital and labor, with the capital share equal to α.

The rate of economic depreciation of noncorporate capital is δ_{2m}. For tax purposes, total depreciation is $\hat{\delta}_{2m}K_{2m,t} + \hat{\delta}_{2x}X_{2m,t}$. We assume that intangible capital investment in the noncorporate sector is negligible and therefore do not include it.[2]

Policy in this economy is a set of tax rates and transfer functions that depend on the state s_t. Both the households and the firms pay taxes.

[2] Most of the investment in the noncorporate sector is in the household and government sectors, with little or no expenditures on items such as research and development, advertising, and investments in organizational capital.

We consider recursive competitive equilibria with equilibrium elements that are stationary functions of the economy's state vector. Because of our assumption that s_t is a Markov process with time-invariant transition probabilities, the aggregate state in period t is $(K_{1m,t}, K_{1u,t}, \hat{K}_{1m,t}, K_{2m,t}, \hat{K}_{2m,t}, s_t)$. For convenience, let $K_t = (K_{1m,t}, K_{1u,t}, \hat{K}_{1m,t}, K_{2m,t}, \hat{K}_{2m,t})$. For a stationary recursive equilibrium, the state in period t is a function of the period t event history $s^t = (s_0,\ldots s_t)$, a fact that we use later.

The problem of the household is to maximize (1) subject to the period t budget constraints:

(2)

$$\begin{aligned}&(1+\tau_{c,t})C_t + A_{t+1}\\&= (1-\tau_{d,t})D_{1,t} + D_{2,t} + (1-\tau_{n,t})W_t L_t + (1+i_t)A_t + T_t,\end{aligned}$$

where A_t is asset holdings at the beginning of period t. The household, during period t, receives income from corporate and noncorporate distributions, $D_{1,t}$ and $D_{2,t}$, respectively, wages at after-tax rate $(1 - \tau_{n,t})W_t$, assets at after-tax rate i_t, and net transfers from the government, T_t. Distributions $D_{1,t}$ and $D_{2,t}$ are both net of taxes paid by corporate and noncorporate firms.

Corporate firms maximize the present value of after-tax distributions to the household, namely,

$$\sum_t \sum_{s^t} (1-\tau_{d,t}(s^t))p_t(s^t)D_{1,t}(s^t).$$

Corporate distributions are

$$\begin{aligned}D_{1,t} = {} & p_{1,t}Y_{1,t} - W_t L_{1,t} - X_{1m,t} - q_t X_{1u,t}\\&-\tau_{1,t}[p_{1,t}Y_{1,t} - W_t L_{1,t} - \hat{\delta}_{1m}\hat{K}_{1m,t} - \hat{\delta}_{1x}X_{1m,t}\\&-\tau_{1k,t}K_{1m,t} - q_t X_{1u,t}]\\&-\tau_{1k,t}K_{1m,t} + \tau_{x,t}X_{1m,t}.\end{aligned}$$

Noncorporate firms maximize the present value of distributions, namely,

$$\sum_t \sum_{s^t} p_t(s^t)D_{2,t}(s^t).$$

Thus, noncorporate distributions are

$$\begin{aligned}D_{2,t} = {} & p_{2,t}Y_{2,t} - W_t L_{2,t} - X_{2m,t}\\&- \tau_{2,t}[p_{2,t}Y_{2,t} - W_t L_{2,t} - \hat{\delta}_{2m}\hat{K}_{2m,t}\\&- \hat{\delta}_{2x}X_{2m,t} - \tau_{2k,t}K_{2m,t}]\\&- \tau_{2k,t}K_{2m,t} + \tau_{x,t}X_{2m,t}.\end{aligned}$$

Note that income taxes are paid once on noncorporate income (net of proprietors' implicit labor income).

There is a final goods producer that combines corporate and noncorporate goods to solve

$$\max_{y_1,y_2} \; F(Y_1,Y_2) - p_1 Y_1 - p_2 Y_2 .$$

The composite output $Y = F(Y_1,Y_2)$ good is used for consumptions and investments:

$$Y_t = C_t + X_{1m,t} + q_t X_{1u,t} + X_{2m,t} + G_t ,$$

where G_t is government consumption. The function F displays constant returns to scale. An implication is that equilibrium distributions are zero, and therefore we do not consider these distributions.

There is growth in the economy due to population growth and productivity. We detrend income and product variables by dividing first by population and second by $(1 + \gamma)^t$, the trend in productivity. To construct hours per capita, we divide total labor input by population. We adopt the notation of lowercase letters for variables that are stationary. For example, $c_t = C_t/[N_t\,(1 + \gamma)^t]$ is detrended consumption and $\ell_t = L_t/N_t$ is detrended labor supply.

It is also convenient to introduce notation for the marginal products of capital. Let $r_{1m,t}$, $r_{1u,t}$, and $r_{2m,t}$ be the marginal products of measured corporate capital, unmeasured corporate capital, and measured noncorporate capital, respectively.

We now are ready to lay out the national accounts for our model economy.

NATIONAL ACCOUNTS TO MODEL ECONOMY ACCOUNTS

In this section, we specify the mapping from U.S. national accounts to the model economy accounts. Our model accounts are summarized in Table A1 in the appendix, with formula specifying entries as a function of model variables. Table A2 reports the main categories of the U.S. national accounts, with average values relative to GDP in the 1990s. Table A3 specifies the model account numbers as a function of the statistics in the U.S. national accounts and the values of these statistics.[3]

Adjustments

There are four important differences between the accounts our model economy dictates and the U.S. national accounts. First, our model output does not include consumption taxes as part of consumption and as part of value added, but NIPA GDP does. A consequence of this is that, unlike NIPA, our accounts are consistent in using producer prices for inputs and outputs. Second, we treat some financial services included in NIPA as intermediate rather than as final. Third, our model treats expenditures on all fixed assets as investment. Thus, consumer durables are treated as an investment in the model accounts rather than as consumption expenditures. We introduce a consumer durable services sector in much the same way as an owner-occupied housing sector is introduced into NIPA. Households rent the consumer durables to themselves. A related adjustment is made for government capital. Finally, and most importantly for the purposes of this paper, our model output includes corporate intangible investment. Intangible investments are expensed and therefore not included in the national accounts.

Adjustments for Consumption Tax. Our *consumption taxes* are all *non-property* taxes on production and imports less subsidies plus business current transfer payments. The reason that we include business transfers in consumption taxes is that they are mostly liability payments, which de facto are a tax. NIPA reports total consumption taxes, which we must assign to the corporate and noncorporate sectors. The sums of consumption and property taxes are reported by sector. We assign aggregate consumption taxes to sectors in proportion to their sums of consumption and property taxes. We subtract the consumption tax from the value added of each sector.

On the product side we assume that all components of NIPA personal consumption expenditures, which include consumer durable expenditures, are taxed at an equal rate. In fact a small part of what we call consumption taxes falls on other components of product, but we do not have good estimates of how much. Thus, we make the simplifying assumption that all consumption taxes fall on personal consumption expenditures. Fortunately, the assignment does not affect the results of this study.

The results of our adjustments for consumption taxes are summarized in Table A3 (lines 4, 8, 10, and 14).

Adjustments for Intermediate Services. Corporate value added includes some services that are not included in our notion of final goods or services.[4] In particular, NIPA imputes to net interest and to consumption an amount equal to the expenses of handling life insurance and pensions, which are intermediate goods in the production of a final good, namely, a diversified financial portfolio. On the income side, we subtract these expenses (which are about 1 percent of GDP) from corporate net interest (Table A2, line 6). On the product side, we subtract these expenses from personal consumption expenditures (Table A2, line 16).

In our mapping from national accounts to model accounts in Table A3, these calculations are listed as "imputed personal business expense." They appear under capital income (line 4) and private consumption (line 10).

Adjustments for Capital Services. We make adjustments in our model accounts for consumer durables and government capital so as to treat them like all other fixed assets accounted for in NIPA.

The implicit rental price of consumer durables that we use is consumer durable depreciation divided by the value of the stock of durables plus the after-tax return on capital. Using estimates from McGrattan and Prescott (2004), we assume that the return is 4.1 percent per year.[5] The imputation to consumption is this rental price times the stock of consumer durables. There are two imputa-

[3] Numbered lines in Table A3 correspond to numbered lines in Table A1.

[4] There have been major changes recently in the accounting of financial services. Most intermediate services are now excluded from GDP.

[5] If i is not close to this value, then returns to capital in the model are not consistent with observed capital stocks and tax rates.

tions to value added. First, we add depreciation of consumer durables to noncorporate depreciation. Second, we add the return on capital times the stock of consumer durables to noncorporate profits.

In Table A3, these calculations are summarized in imputed capital services under noncorporate capital income (line 8) and private consumption (line 10).

In NIPA, the services of government capital are equal to its depreciation. Thus, net income is zero. We define net income of government capital as our average after-tax return on capital (4.1 percent) times the value of this capital stock. We add this income to profits of the noncorporate sector on the value added side of the accounts and to public consumption on the product side.

In Table A3, these calculations are summarized in imputed capital services under noncorporate capital income (line 8) and public consumption (line 11).

Adjustments for Intangible Investments. The last adjustment we make to corporate income is to add back investments that had been expensed. We do not have direct measures of these expenses but can infer them from our theory and NIPA data. In this section, we use the theory to estimate the average level of intangible investment and the equilibrium path since 1990.

The Average Level of Intangible Investment in the 1990s. As in McGrattan and Prescott (2004), we can take an indirect approach, using observations on corporate profits and returns to tangible assets to estimate a return to intangible assets. NIPA profit before corporate income tax is

$$\begin{aligned}&\text{NIPA profit} =\\ &r_{1m}k_{1m} + r_{1u}k_{1u} - \delta_{1m}k_{1m} - \tau_{1k}k_{1m} - qx_{1u}\\ &= (r_{1m} - \delta_{1m} - \tau_{1k})k_{1m} + q(r_{1u}/qk_{1u} - x_{1u}).\end{aligned} \tag{3}$$

If economic and accounting depreciation are equal and returns are equated for all assets, then the first-order conditions of the model in the section "The Model Economy" imply that the following hold[6]:

$$i = r_{1u}/q - \delta_{1u} \tag{4}$$

$$i = \frac{(1-\tau_1)(r_{1m} - \tau_{1k}) + \tau_1\delta_{1m}}{1-\tau_x} - \delta_{1m} \tag{5}$$

$$x_{1u} = (\gamma + \eta + \delta_{1u})k_{1u} \tag{6}$$

$$x_{1m} = (\gamma + \eta + \delta_{1m})k_{1m} \tag{7}$$

on a balanced growth path, where i is the real interest rate and η is the population growth rate.

Equations (3) through (7) can be solved for the average level of intangible investment and capital. This is done as follows. We use BEA data to get estimates of the corporate income tax rate, τ_1, the corporate property tax rate, τ_{1k}, the subsidy to investment, τ_x, the tangible depreciation rate, δ_{1m}, and corporate tangible investment, x_{1m}. We can use either the noncorporate returns or estimates of preference parameters to get the real interest rate, i. Population growth, η is around 1 percent per year. Trend technology growth, γ, is around 2 percent per year.

The system of equations (3) through (7) is five equations in the five unknowns, $r_{1u}/q - \delta_{1u}$, r_{1m}, qk_{1u}, k_{1m}, and qx_{1u}. Using data from the 1990s, our estimate of the average value of intangible capital, qk_{1u}, is 0.65 times GDP.[7] This estimate is independent of our choice of δ_{1u}. Our estimate of net investment, $qx_{1u} - \delta_{1u}k_{1u}$, is also independent of our choice of δ_{1u}. Net intangible investment averaged 2 percent of GDP.[8]

The Equilibrium Path of Intangible Investment Since 1990. We can infer the path for intangible investment using intratemporal first-order conditions of the model. We use two approaches that lead to similar quantitative implications for productivity in the 1990s. The two approaches rely on different assumptions about cost shares over the business cycle.

Our first approach assumes that the capital income share does not vary over the cycle. Let θ be the capital share in the corporate sector, which

[6] We also did calculations allowing for differences in the depreciation rates. Holding the corporate tax rate fixed, we find a higher average level of intangible investment if accounting depreciation exceeds economic depreciation. Thus, to be conservative in our conclusions about the importance of intangible investment, we assume economic and accounting depreciation are equal.

[7] The inputs to this calculation are $\tau_1 = 0.37$, $\tau_{1k} = 0.02$, $\tau_x = 0$, $\delta_{1m} = 0.06$, $x_{1m} = 0.099$, $i = 0.041$, $\gamma = 0.02$, and $\eta = 0.01$.

[8] For convenience we will set $\delta_{1u} = 0$ when we derive time series for $q_t x_{1u,t}$. We are, in effect, working with net intangible investment.

Figure 1

Intangible Investment (Corporate Income Shares Fixed)

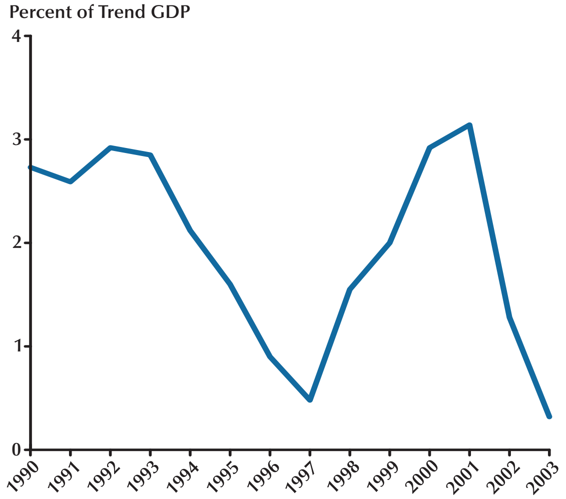
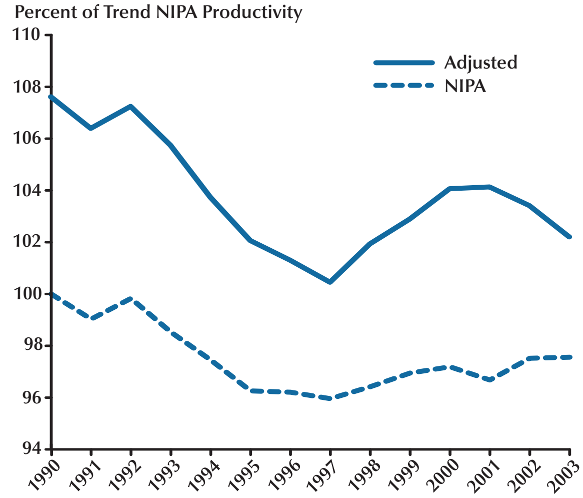


Figure 2

Productivity Relative to a 2 Percent Trend (Corporate Income Shares Fixed)



we take to be the average corporate capital income share, $(r_{1m}k_{1m} + r_{1u}k_{1u})/(p_1y_1)$, or, equivalently, 1 less the average corporate labor income share, $1 - w\ell_1/(p_1y_1)$. Using averages in the 1990s, we estimate an average corporate capital share of $\theta = 0.33$. If intangible investments are chosen so that

$$\frac{w_t\ell_{1,t}}{p_{1,t}y_{1,t}} = 1-\theta_t$$
$$= 1-\theta,$$

then $q_tx_{1u,t}$ satisfies

$$w_t\ell_{1,t} = (1-\theta)p_{1,t}y_{1,t}$$
$$= (1-\theta)[\mathrm{va}_{1,t}^{\mathrm{accounting}} + q_tx_{1u,t}], \tag{8}$$

where we have an estimate of θ and time series for corporate compensation ($w_t\ell_{1,t}$) and accounting corporate value added ($\mathrm{va}_{1,t}^{\mathrm{accounting}}$). The unknown in equation (8) is the product $q_tx_{1u,t}$.

In Figure 1, we display the implied time series for intangible investment after 1990. What is striking about this figure is the sixfold increase in the level of intangible investments between 1997 and 2000. This represents a very large change in investment. This change in investment has consequences for output, profits, and total investment.

Output is the sum of corporate and noncorporate income, namely, y_t, after the relevant adjustments to the national accounts are made. Economic corporate profits are capital income, which is corporate income less labor income and depreciation. If we assume an intangible depreciation rate of zero, then economic corporate profits are given by[9]

$$p_{1,t}y_{1,t} - w_t\ell_{1,t} - \delta_{1m}k_{1m,t}.$$

Economic corporate investment is the sum of tangible plus intangible investments. Economic profit shares and investment shares are defined relative to output, y_t, rather than GDP.

In Figure 2, we compare the standard measure of productivity, real GDP per hour worked, with our economic measure y_t/ℓ_t. The hours measure we use is described in Prescott and Ueberfeldt (2003) and based on the Current Population Survey. We normalize hours to be 1 in 1990 so that we can directly compare the magnitudes of

[9] In McGrattan and Prescott (2004), we do a sensitivity analysis that includes varying δ_{1u}.

Figure 3

Corporate Profit Shares of Income (Corporate Income Shares Fixed)

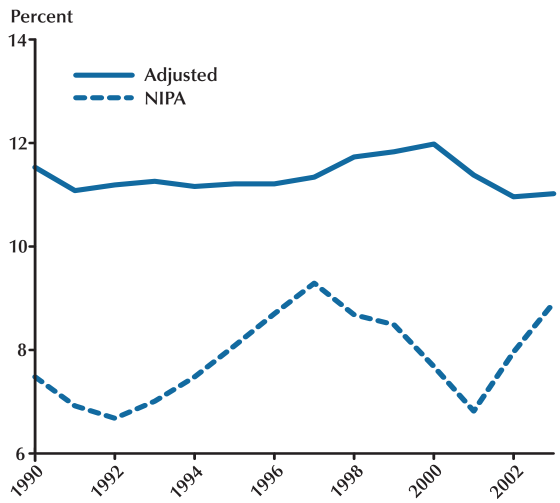
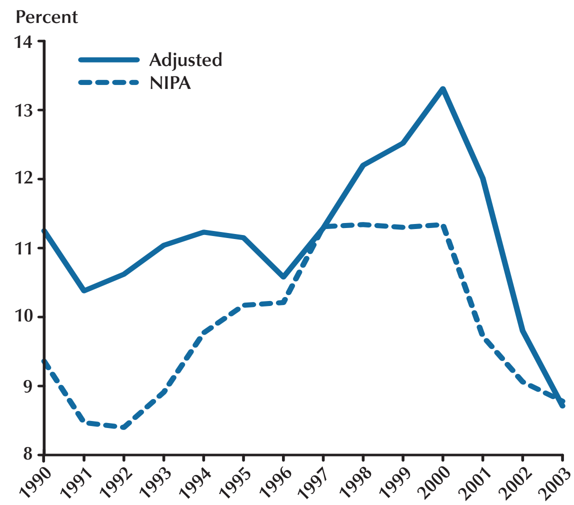


Figure 4

Corporate Investment Shares of Income (Corporate Income Shares Fixed)



GDP and y. We also divide both measures by 1.02^t, the historical trend.

The figure shows that economic productivity fell faster than accounting productivity in the early 1990s but grew much faster at the end of the 1990s. Notice that economic productivity is higher than accounting productivity in 1990 because output, y, is 8 percent higher than GDP. A comparison with Figure 1 illustrates how important the movements in intangible investment are for output.

In Figure 3, we plot the accounting measure of corporate profits relative to GDP and our economic corporate profits relative to output. Our measure is significantly higher because we do not subtract intangible investment or property tax. Our measure also includes the small part of corporate net interest that is not intermediate services. Our profit share is about 12 percent, whereas NIPA's profit share is 7.5 percent.

Of particular significance is the fact that the patterns are very different. In the late 1990s, economic profits are high, while NIPA profits are low.

In Figure 4, we plot the accounting measure of the corporate investment share, namely, corporate tangible investment relative to GDP and our economic measure, which is total corporate investment—tangible and intangible—divided by output, y. Notice that the standard accounting measure shows only a modest investment boom in the late 1990s, whereas our measure shows a bigger investment boom.

Our second approach to measuring the path of intangible investment assumes that the ratio of labor income shares across sectors is constant. Let $1-\theta_t$ be the corporate labor income share in period t and let $1-\alpha_t$ be the noncorporate labor income share in period t. Thus, our second approach assumes that

$$\frac{1-\theta_t}{1-\alpha_t}=\frac{1-\theta}{1-\alpha}, \tag{9}$$

where θ and α are found by taking averages over our sample period. For the corporate sector, the average is $\theta = 0.33$. For the noncorporate sector, the average is $\alpha = 0.496$.[10]

Assuming that corporate income shares stay constant puts a lower bound on the increase of intangible investment during the late 1990s, as

[10] The capital cost share for the noncorporate sector is high because a significant fraction of this sector's capital is housing and consumer durables, which have a capital cost share near 1.

Figure 5

Intangible Investment (Corporate Income Shares Varying)

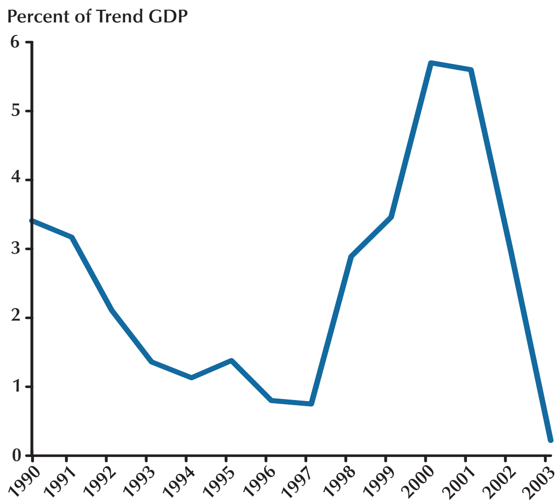
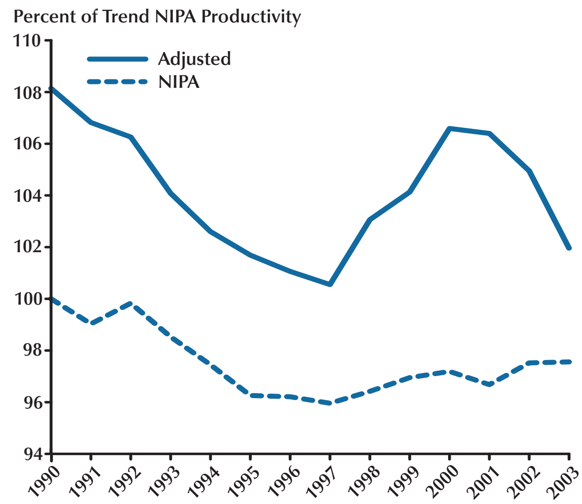


Figure 6

Productivity Relative to a 2 Percent Trend (Corporate Income Shares Varying)



capital income shares are almost surely procyclical. The reason is simple. If accounting profits are low relative to trend, we are attributing the difference to expensed investments. However, accounting profits may appear even lower if compared with boom-time levels. When we assume that income shares vary over the cycle, then we find a larger gap between economic and accounting profits in booms and, hence, a larger amount of intangible investment.

Equation (9) and observables can be used to find the value of intangible investment in units of the consumption good as follows. Substituting for the labor shares in (9) yields

$$\left(\frac{w_t \ell_{1,t}}{p_{1,t} y_{1,t}}\right)\left(\frac{p_{2,t} y_{2,t}}{w_t \ell_{2,t}}\right) = \frac{1-\theta}{1-\alpha}.$$

This equation can be solved for corporate value added, $p_{1,t}y_{1,t}$, as a function of observables. Variable $p_{1,t}y_{1,t}$ can be used along with accounting value added in the corporate sector to find the value of unmeasured investment, that is,

$$q_t x_{1u,t} = p_{1,t} y_{1,t} - \text{va}_{1,t}^{\text{accounting}}.$$

In Figure 5, we plot the equilibrium path for the implied investment in intangibles. As before, we find a sixfold increase in the level of intangible investments between 1997 and 2000. The main difference between the measures in Figure 1 and Figure 5 are the magnitudes. Assuming varying income shares implies a higher absolute value of intangible investment at the peak in 2000.

What are the consequences for productivity? In Figure 6, we compare the standard measure of productivity, real GDP per hour worked, with the economic measure y_t/ℓ_t adjusted for the intangible investment in Figure 5. Again, we normalize hours to be 1 in 1990 so that we can directly compare the magnitudes of GDP and y. We also divide both measures by 1.02^t, the historical trend.

As in the case of fixed corporate shares, we find that economic productivity fell faster than accounting productivity in the early 1990s but grew much faster at the end of the 1990s. The rise in productivity is somewhat higher in the case where corporate income shares vary. Over the period 1997-2000, we estimate that productivity rose 3.2 percent per year in the case with fixed corporate income shares and 4 percent per year in the case with varying corporate income shares.

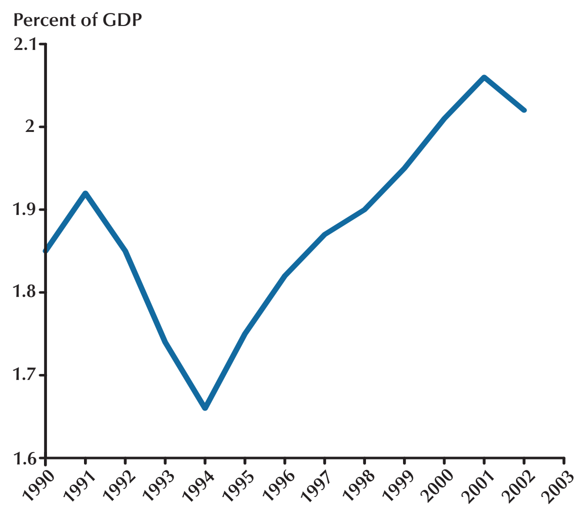
However, both cases show a much different picture than GDP per hour.

SUPPORTING EVIDENCE

As we noted earlier, we do not have direct measures of all intangible investments. But there are some direct measures of one important component of intangible investment, namely, research and development. In Figure 7, we plot expenditures for research and development performed by industry. Some of these expenditures are capital expenditures and therefore are not included in our notion of intangible investments. However, we find a similar pattern of investment. Investment in research and development fell rapidly in the first half of the 1990s and rose rapidly in the second half. This is what we find for total intangible investment.

Figure 7

Industry Research and Development



SUMMARY

U.S. growth in GDP per hour worked, which we call accounting productivity, was well below trend in the 1973-95 period and then recovered to the historical level of 2 percent per year growth beginning in 1995. (See Figure 2.) This picture is what most of the leading researchers in productivity accounting find.[11] They, as do we, use hours worked estimates based on the Current Population Survey.[12]

We find that economic productivity, which includes corporate intangible investment in output, displays a very different pattern. As shown in Figure 2, we find that there was a productivity growth boom in the late 1990s with productivity growth well in excess of the 2 percent historical trend. Prior and subsequent to this boom, average productivity growth was about half of the level of trend growth.

Our accounting resolves the puzzle of why corporate accounting profits were so low in the late 1990s boom. Corporations were making large intangible investments, which lowered their accounting profits, but not their economic profits. The economic profits share of economic income was high in the boom. (See Figure 3.) Our accounting also resolves the puzzle of why investment share of output was not much higher in this boom than standard accounting figures indicate. With the accounting numbers dictated by economic theory, the share increases, and increases a lot in the boom. (See Figure 4.)

[11] See, for example, Jorgenson, Ho, and Stiroh (2003, p. 45) and Groningen Growth and Development Centre and The Conference Board (2004).

[12] There are two other measures of U.S. hours worked. The BEA estimate, which uses state unemployment tax records to estimate the number of paid workers, and the BLS estimate, which uses an establishment survey to estimate the number of paid workers. The BEA hours estimate, which is an annual series, paints the same picture as CPS hours. The BLS hours estimate, which is a quarterly series, paints a different picture in the post-2000 period. With the BLS estimate of aggregate hours, estimated productivity growth is 1 percent higher in the 2000-03 period. See Kunze (2004) and Eldridge, Manser, and Otto (2004) for details.

REFERENCES

Board of Governors of the Federal Reserve System (1945-2004). *Flow of Funds Accounts of the United States*. Washington, DC: Federal Reserve System.

Eldridge, Lucy P.; Manser, Marilyn E. and Otto, Phyllis Flohr. "Alternative Measures of Supervisory

Employee Hours and Productivity Growth." *Monthly Labor Review*, April 2004, *127*(4), pp. 9-28.

Groningen Growth and Development Centre and The Conference Board. *Total Economy Database*, 2004; www.ggdc.net.

Jorgenson, Dale W.; Ho, Mun S. and Stiroh, Kevin J. "Lessons for Europe from the U.S. Growth Resurgence." *CESifo Economic Studies*, 2003, *49*(1), pp. 27-47.

Kunze, Kurt. "U.S. Hours Worked Data." Memo, Bureau of Economic Analysis, August 2004.

McGrattan, Ellen R. and Prescott, Edward C. "Taxes, Regulations, and the Value of U.S. and U.K. Corporations." Staff Report 309, Federal Reserve Bank of Minneapolis, 2004 (forthcoming in *Review of Economic Studies*).

Prescott, Edward C. and Ueberfeldt, Alexander. "U.S. Hours and Productivity Behavior Using CPS Hours Worked Data: 1959-I to 2003-II." Unpublished manuscript, Federal Reserve Bank of Minneapolis, 2003.

U.S. Department of Commerce, Bureau of Economic Analysis (1929-2004). *Survey of Current Business*. Washington, DC: U.S. Government Printing Office.

APPENDIX

In this appendix, we display the national accounts that we work with, the NIPA categories before we make our adjustments, and a mapping from the national accounts to the model accounts. Table A1 lists the account categories for our model along with formulas for variables in the model. Table A2 lists the NIPA categories along with average values relative to GDP in the 1990s, which give the reader a sense of the magnitudes of the adjustments. Table A3 provides a mapping between these accounts. In the main text, we provide justifications for the calculations summarized here.

Table A1

Model Economy Accounts

		Model expression
1	**Corporate domestic value added**	p_1y_1
2	Depreciation	$\delta_{1m}k_{1m}$
3	Labor income	$w\ell_1$
4	Capital income	$r_{1m}k_{1m} + r_{1u}k_{1u} - \delta_{1m}k_{1m}$
5	**Noncorporate domestic value added**	p_2y_2
6	Depreciation	$\delta_{2m}k_{2m}$
7	Labor income	$w\ell_2$
8	Capital income	$r_{2m}k_{2m} - \delta_{2m}k_{2m}$
9	**Total domestic value added**	y
	Domestic Product	
10	Private consumption	c
11	Public consumption	g
12	Corporate measured investment	x_{1m}
13	Corporate unmeasured investment	qx_{1u}
14	Noncorporate investment	x_{2m}
15	**Total domestic product**	y

Table A2

National Accounts, Average in 1990s Relative to GDP

1	**Corporate domestic value added**	0.589
2	Consumption of fixed capital	0.066
3	Compensation of employees	0.378
4	Corporate profits with IVA and CCadj	0.075
5	Taxes on production and imports[1]	0.056
6	Net interest and miscellaneous payments	0.013
7	**Noncorporate domestic value added**	0.400
8	Consumption of fixed capital	0.053
9	Compensation of employees	0.240
10	Rental income of persons with IVA	0.014
11	Proprietors' income with IVA and CCadj	0.068
12	Taxes on production and imports[1]	0.020
13	Net interest and miscellaneous payments	0.051
14	Statistical discrepancy	0.011
15	**Total domestic value added**	1.000
	Domestic Product	
16	Personal consumption expenditures	0.670
17	Durable goods	0.082
18	Nondurable goods and services	0.588
19	Government consumption expenditures and gross investment	0.189
20	Consumption expenditures	0.156
21	Gross investment	0.033
22	Gross private domestic investment[2]	
23	Corporate	0.099
24	Noncorporate	0.055
25	Net exports of goods and services	–0.013
26	**Total domestic product**	1.000
	Addendum	
27	Consumption taxes	0.047
28	Consumption of fixed capital, durable goods	0.062
29	Current-cost net stock of government fixed assets	0.604
30	Current-cost net stock of consumer durable goods	0.308

NOTE: IVA, inventory valuation adjustment; CCadj, capital consumption adjustment.

[1] This category includes business transfers and excludes subsidies.

[2] The breakdown into corporate and noncorporate investments is based on data from the *Flow of Funds Accounts* (Board of Governors of the Federal Reserve System, 1945-2004). For corporate investment, we sum investment of nonfinancial corporate business, financial corporations, and 10 percent of farm business.

Table A3

Mapping from National Accounts to Model Accounts[1]

		Model	NIPA
1	**Corporate domestic value added** (p_1y_1)		
2	Depreciation ($\delta_{1m}k_{1m}$)	0.066	
	Consumption of fixed capital		0.066
3	Labor income ($w\ell_1$)	0.378	
	Compensation		0.378
4	Capital income ($r_{1m}k_{1m} + r_{1u}k_{1u} - \delta_{1m}k_{1m}$)	0.120	
	Corporate profits with IVA and CCadj		0.075
	Net interest and miscellaneous payments		0.013
	Less: Imputed personal business expense[2]		–0.010
	Taxes on production and imports		0.056
	Less: Consumption taxes		–0.034
	Corporate unmeasured investment		0.020
			0.120
5	**Noncorporate domestic value added** (p_2y_2)		
6	Depreciation ($\delta_{2m}k_{2m}$)	0.115	
	Consumption of fixed capital		0.053
	Consumption of fixed capital, durable goods		0.062
			0.115
7	Labor income ($w\ell_2$)	0.251	
	Compensation		0.192
	70% Proprietors' income with IVA and CCadj		0.048
	Statistical discrepancy		0.011
			0.251
8	Capital income ($r_{2m}k_{2m} - \delta_{2m}k_{2m}$)	0.132	
	Rental income of persons with CCadj		0.014
	30% Proprietors' income with IVA and CCadj		0.020
	Net interest and miscellaneous payments		0.051
	Current surplus of government enterprises		0.001
	Taxes on production and imports		0.020
	Imputed capital services[3]		0.038
	Less: Consumption taxes		–0.012
			0.132
9	**Total domestic value added** (y)	1.062	1.062

Table A3 cont'd

		Model	NIPA
	Domestic product		
10	Private consumption (c)	0.611	
	Personal consumption expenditures		0.670
	Less: Consumption taxes		–0.042
	Imputed capital services[3]		0.013
	Consumption of fixed capital, durable goods		0.062
	Less: Consumption expenditures, durable goods		–0.082
	Less: Imputed personal business expense[2]		0.010
			0.611
11	Public consumption (g)	0.180	
	Government consumption expenditures		0.156
	Imputed capital services[3]		0.025
			0.180
12	Corporate measured investment (x_{1m})	0.099	
	Gross domestic private investment, corporate		0.099
13	Corporate unmeasured investment (qx_{1u})	0.020	
	Corporate unmeasured investment		0.020
14	Noncorporate investment (x_{2m})	0.152	
	Gross domestic private insurance, noncorporate		0.055
	Personal consumption expenditures, durable goods		0.082
	Less: Consumption taxes		–0.005
	Government investment		0.033
	Net exports		–0.013
			0.152
15	**Total product** (y)	1.062	1.062

NOTE: [1] Model and NIPA values based on averages over 1990s in Table A2.

[2] Expense is for handling life insurance and pension plans.

[3] Imputed capital services are equal to 4.1 percent times the current-cost net stock of government fixed assets and consumer durable goods.

Commentary

Ricardo J. Caballero

Ellen McGrattan and Ed Prescott's article is right in highlighting an important fact: "Intangible investment" is large and likely to vary over time. Unfortunately, standard methods for calculating national income and product accounts miss the bulk of this investment. Thus, conclusions about investment, productivity growth, and corporate sector performance can be severely distorted.

The authors go on to make a bold claim: They can measure intangible investment correctly with the help of a simple theory.

The result of their bold attempt is to change the "history" of the late 1990s—as, in their preferred version, productivity growth and other prosperity indicators look significantly better than those obtained with conventional national accounts data.

I like bold statements. They are a useful starting point and they invite reactions. My comments below start the latter. The current version of the paper incorporates my main suggestion during the conference, which was to provide a measure of the path of intangible investment that does not depend on verifiable assumptions that are rejected in the data. This is now their first approach. But they also chose to keep their previous method in the paper (now called the second approach) and much of the wording in the introduction and conclusion fits the second approach better than the first. Thus, I have decided to revisit the comments I made at the conference.

I had essentially two points: a minor one on the authors' characterization of the 1990s and a more substantive one on their measurement of intangible capital.

ON THE 1990S

In the abstract, the paper states that "ignoring intangible investments in the late 1990s leads one to conclude that productivity growth was modest, corporate profits were low, and corporate investment was at moderate levels."

This gloomy characterization of the (incorrectly measured) 1990s does not fit conventional wisdom and the (incorrect) facts. It may well be the case that correctly accounting for intangible investments makes things look better; but, even without this correction, the late 1990s are perceived as a time of massive investment, fast productivity growth, and corporate bonanza.

By 1995-2000, productivity growth had returned to levels not seen since the late 1960s, and in high-technology sectors, such as industrial machinery and electronic machinery, productivity growth reached astonishing numbers—well over 5 percent per year (see Table 1).

Similarly, investment as a share of gross domestic product reached record levels, especially when measured as a ratio of real quantities, as the rate of decline in the relative price of equipment accelerated during the late 1990s. Finally, while profits may not have accelerated in tandem with investment and productivity, capital owners did extremely well as the effective cost of capital declined and capital gains increased dramatically.

ON THE MODEL AND MEASUREMENT

Here one needs to differentiate between comments on the many steps to improve the

Ricardo J. Caballero is a Ford International Professor of Economics at the Massachusetts Institute of Technology.

Federal Reserve Bank of St. Louis *Review*, July/August 2005, *87*(4), pp. 551-53.

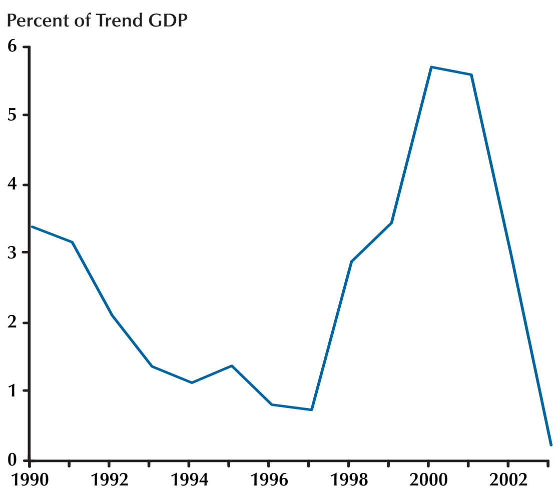
Table 1

Private Nonfarm Business Productivity Growth

	1948-73	1973-79	1979-90	1990-95	1995-2000
Labor productivity	2.9	1.2	1.4	1.6	**2.5**
Multifactor productivity	1.9	0.4	0.3	0.6	**1.1**
Manufacturing	1.5	–0.6	1.1	1.3	**2.1**
Industrial machinery	0.7	0.2	3.2	3.1	**5.8**
Electronic machinery	2.1	1	3	6	**7.4**

Figure 1

Intangible Investment



accounting of averages and comments on the dynamics. I have little to say about the former, except to express admiration for the many useful and careful steps taken in "amending" conventional national accounts procedures. My concerns are all about the discussion of dynamics.

On the investment side, the authors assume no adjustment costs and, hence, a marginal product of capital equal to the interest rate throughout. This is not a good model of short-run investment. Short-run frictions are of the essence in investment theory, and the corresponding capital gains can generate large wedges between marginal product and interest rates. Because several of the numbers computed in the paper are the result of ratios of numbers that are very close to zero, it is easy to generate variations of several hundred percent in the numbers reported in the paper, just by allowing for small wedges between marginal product and interest rates.

On the labor side, there are assumptions of perfect labor mobility across sectors, Cobb-Douglas production functions, and so on. Again, while these may be reasonable assumptions for the medium and long run, they are inadequate for the short run. This difficulty is particularly apparent in their main expression used in the conference to measure the path of intangible capital in the corporate sector:

$$(1) \qquad x_1 = \frac{(1-\alpha)}{(1-\theta)}\left(\frac{CCOMP}{NCOMP}\right)NVA - CVA,$$

where x is intangible investment; *CCOMP* and *NCOMP* are labor compensation in the corporate and noncorporate sectors, respectively; *NVA* and *CVA* are value added in the corporate and noncorporate sectors, respectively; and θ and α are the Cobb-Douglas share of labor in the corporate and noncorporate sectors, respectively. This leads to Figure 5 in the current version of their paper (my Figure 1).

This figure produces a dramatic rise in intangible investment at the end of the 1990s and is the figure that best matches the introduction and general message of the paper. But how did they obtain equation (1), and how important are the assumptions behind this derivation for the results? The answer to the second part of the question is "a lot." Let me follow a slightly different derivation from theirs, which facilitates understanding why this is so.

Figure 2

Intangible Estimates for Different "1" Operations

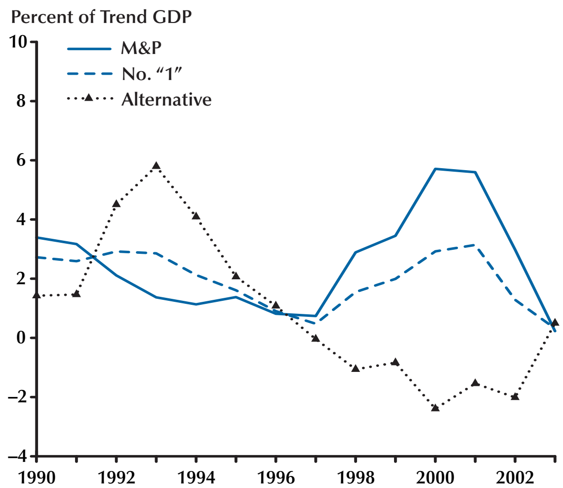
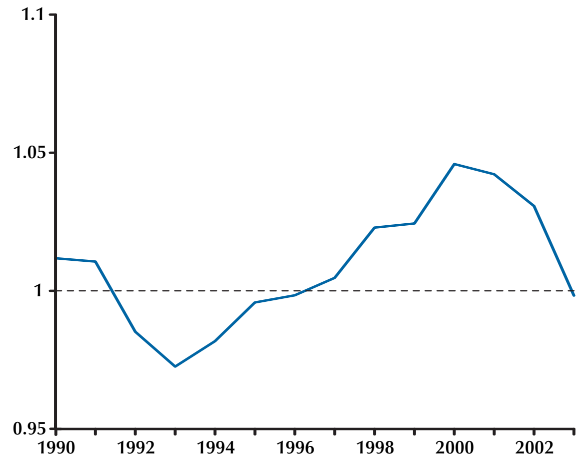


Figure 3

"1"



Following what they now refer to as "the first approach," one can obtain an expression for intangible capital from the first-order condition of the corporate sector, after assuming a Cobb-Douglas production function for actual (as opposed to measured) value added, and noticing that actual value added is equal to measured value added plus intangible capital (erroneously expensed). Then,

$$x_1 = CCOMP/(1-\theta) - CVA. \quad (2)$$

But how do we go from equation (2) to (1)? Before answering this, note that under a Cobb-Douglas assumption in the noncorporate sector, the basic first-order condition for labor implies:

$$\text{“1”} = (1-\alpha)NVA/NCOMP. \quad (3)$$

We are now ready to go from (2) to (1), as the latter is obtained by multiplying the first term on the right-hand side of (2) by the expression for "1."

However, the nice feature of 1 is that we can do many things with it without affecting the expression we are multiplying or dividing it by. In particular, we may chose not to multiply by "1," or we can chose to divide the first expression of the right-hand side of (2) and multiply the second expression. All of these formulas for intangible capital should yield the same result. My Figure 2 shows that they do not.

The dashed line corresponds to my suggestion at the conference and their current Figure 1; the solid line corresponds to their Figure 4; and the remaining line corresponds to the other transformation. It is apparent from this figure that their "1" is not really 1. This is confirmed in Figure 3.

In summary, I maintain my recommendation from the conference. Given that "1" is not really 1, I would suggest they focus on the measure of intangible investment that does not use this incorrect information. This is the dashed line in my Figure 2. Of course, in this case the story for the end of the 1990s is much less dramatic, and it looks more like a story about the dip of intangible investment during the mid-1990s rather than a surge at the end of the 1990s. Where would this dip come from? There is a good chance that the implicit "1" in the corporate sector (also Cobb-Douglas assumption, no frictions, etc.) is not correct either, in which case it may all be just "measurement" error.

Having said this, I believe the step taken by the authors was worth it. It was bold, most likely not right, but it opens a potentially important area of research. Somebody will get it right in the future. The authors will then get the credit they deserve.

Organizational Dynamics Over the Business Cycle: A View on Jobless Recoveries

Kathryn Koenders and Richard Rogerson

This paper proposes a new explanation for the apparent slow growth in employment during the past two recoveries. The authors' explanation emphasizes dynamics within growing organizations and the intertemporal substitution of organizational restructuring. A key implication of the analysis is that recoveries from recessions following long expansions will have slower employment growth. Empirical analysis shows that the recovery that began in 1970 also exhibited slow employment growth, consistent with this prediction of the analysis.

Federal Reserve Bank of St. Louis *Review*, July/August 2005, *87*(4), pp. 555-79.

Since the work of Burns and Mitchell (1946), economists who study business cycle fluctuations typically refer to the "business cycle facts" without need to reference a particular episode in a particular country. One of the accepted stylized facts of business cycle movements is that employment and output are strongly positively correlated, although employment lags output by about one quarter. The apparent slow growth of employment in the recoveries following the past two U.S. recessions (i.e., the so-called "jobless recovery" phenomenon) runs counter to this stylized fact. This paper suggests a possible explanation for this apparently anomalous behavior.

The two most recent recessions in the United States share a common property: Both followed unusually long expansions. Motivated by this observation, we propose an economic mechanism that links the speed at which employment increases during the recovery from a recession to the length of the expansion preceding the recession.[1] This mechanism stresses the manner in which organizations seek to eliminate unneeded labor. Specifically, we assume that inefficiencies regarding the use of labor emerge over time within an organization. Eliminating these inefficiencies (a process we refer to as reorganizing) requires scarce organizational resources that must be diverted away from current production. This trade-off generates opportunities for intertemporal substitution, and we show that reorganization will be postponed to periods in which production is relatively low. It follows that after a long expansion, many more organizations have postponed reorganization. Because reorganization leads to the shedding of unnecessary labor and takes time, this gives rise to an extended period in which the economy sheds labor, thereby delaying the date at which aggregate employment begins to increase during the recovery.

The first part of this paper presents one formulation of a model of organizations that generates these effects. The core model should be seen as an extension of the Lucas (1978) span-of-control model and the Hopenhayn (1992) industry equilib-

[1] As noted by Kliesen (2003), these two recessions are also relatively mild. Our theory does not address this regularity.

Kathryn Koenders is a doctoral student and Richard Rogerson is the Rondthaler Professor of Economics at the W.P. Carey School of Economics at Arizona State University. The authors thank Fernando Alvarez and conference participants and seminar participants at the Federal Reserve Banks of Atlanta, Cleveland, New York, and Philadelphia for comments. Richard Rogerson acknowledges support from the National Science Foundation.

rium model to allow for a richer set of dynamics within an organization. Although we focus on the implications for business cycles, we believe this model may also prove useful for examining plant- and firm-level dynamics more generally. The model is purposefully simplified to highlight the key economic trade-offs, and the analysis focuses on the qualitative nature of the interactions. The task of building a model suitable for quantitative analysis of the forces is left for future work.[2]

Any theory that links the speed of the recovery in employment to the length of the preceding expansion would seem to be a viable candidate for explaining the anomalous behavior of employment following the past two recessions. However, because the recession of 1969-70 also followed an unusually long expansion, an obvious implication of any such theory is that a slow recovery of employment should also have been observed following this recession. The second part of the paper turns to this issue and argues that, when viewed from the perspective of our model, the behavior of employment in the 1970 recovery is in fact very similar to the behavior of employment in the recoveries of 1991 and 2001 and is qualitatively different from the behavior of employment in the other post-World War II recoveries.

One interpretation of our explanation is that the recent recessions do not represent counterexamples to the standard set of business cycle facts, but rather that the business cycle facts need to be modified somewhat to acknowledge that recoveries following long expansions exhibit somewhat different dynamics. There are, of course, other types of explanations that one might appeal to. For example, if one thought that it is only the two most recent recessions that appear different, one might consider the possibility that the business cycle facts are evolving and seek to understand what features of the economy are changing that would lead to the change in business cycle dynamics of employment. Schreft and Singh (2003) pursue this tack, arguing that increased flexibility in personnel policies are responsible for the change. A second general class of explanations is to posit that the anomalous behavior of employment in any particular cycle is due to an additional shock or policy change that happens to coincide with the recovery. The work of Cole and Ohanian (2004) on employment during the second half of the Great Depression is exactly this type of explanation: They argued that employment did not recover as one would have expected because of the adoption of the New Deal policies. A third and related class of explanations stresses that the economic changes taking place in the background may influence the nature of cyclical episodes. Along these lines, a commonly heard explanation for the recent slow recovery of employment was that the world had become more uncertain, leading firms to postpone increases in employment. Another explanation in this class is that the current period involves a greater degree of structural change. Groshen and Potter (2003) examine data on sectoral employment shares and argue that this is the case.[3] Andolfatto and MacDonald (2004) present a model in which certain types of technological change can generate this outcome. We do not compare our explanation with these others, but do note that the extent to which the 1970 recovery is viewed as being similar to the two most recent recoveries would cast some doubt on the theory that there is evolution in the business cycle facts.

The mechanism that we describe is related to others that have appeared in the literature, and it is of interest to note the similarities and differences. An old idea in the business cycle literature is that recessions are periods of restructuring. However, modern formulations of this idea, such as Lilien (1982), are based on the notion that the key element of restructuring is across organizations—in particular, that resources need to be reallocated across sectors. In contrast, our model does not stress the reallocation of resources from one organization to another but rather the restructuring that takes place within organizations that leads to the elimination of wasteful employment. Hall (1991) argued that recessions should be

[2] van Rens (2004) presents an alternative theory that stresses investment in organizational capital as a key factor in explaining the pace of employment growth coming out of a recession.

[3] See Aaronson, Rissman, and Sullivan (2004) for an argument against this interpretation.

thought of as reorganizations, but the reorganizations that he stresses were really reallocations of labor across activities. As noted here previously, Groshen and Potter (2003) have stressed that the amount of reallocation needed may vary over time and be higher in some business cycle episodes than in others.

A second related literature is associated with the work of Caballero and Engel (1999) and Krusell and Smith (1998). Both papers consider the possibility that the distribution of individual state variables can influence how the economy responds to shocks. Caballero and Engel argue that the distribution of the difference between plant-level capital stocks and their ideal points was quantitatively significant for the response of the economy to shocks. Krusell and Smith examine how the distribution of asset holdings across consumers affects propagation of shocks. Qualitatively, our model emphasizes a similar channel, because we argue that the distribution of efficiencies of organizations matters for how the economy responds to shocks. However, despite the similarity, the mechanics are quite different. In particular, if carried over to the labor setting, the Caballero and Engel model cannot explain why aggregate demand for labor would continue to decrease in the face of positive aggregate shocks, a result that can emerge in our model and is central to accounting for the delayed increase in employment that accompanied the past two recoveries.

An outline of the next five sections follows: We describe our benchmark model, which captures the evolution of an individual organization over its life cycle. Then we consider how aggregate temporary shocks interact with the decisions of the organization and derive our key result: Organizations will concentrate reorganizations during periods with negative aggregate shocks. The last sections (i) discuss the implications of this finding for the cyclical properties of labor demand and (ii) carry out an empirical analysis of postwar business cycles to show that the recovery following the 1969-70 recession exhibits patterns for employment that are quantitatively very similar to those found in the two most recent recoveries.

BENCHMARK MODEL: A LIFE CYCLE MODEL OF ORGANIZATIONAL DYNAMICS

In this section we formulate a model of the life cycle of an organization. In the next section we will use this benchmark model to investigate how the resulting pattern of organizational dynamics may be affected by shocks that we interpret to be business cycle shocks. The goal of these two sections is to highlight a particular interaction between organizational dynamics and business cycle shocks.[4] With this in mind, we purposefully work in a very simple setting to best highlight this interaction. We leave the development of a model that would be useful for a quantitative assessment of these interactions to future work.

The essence of the benchmark model is as follows. We view an organization as producing a differentiated product and therefore facing a downward sloping demand curve for its product. Our model captures the following stylized evolution of an organization over its life cycle: When an organization is first created it faces a relatively low demand for its product. But, over time, this situation may change. Some organizations fail and disappear, while others experience large increases in the demand for their product and hence grow. However, even those organizations that grow and become large will at some point experience decreases in their demand and eventually cease to exist.

The model that we describe here is one of an individual organization that faces stochastic demand for its product but takes all input prices as given, prices that remain constant over time. We assume that the organization maximizes the present discounted value of profits using a discount rate of β, which one can think of as 1 divided by the sum of 1 plus the interest rate. We now describe the specifics of the model in more detail.

---

[4] Our model without aggregate shocks can also be viewed as extending the model of Hopenhayn (1992) to consider richer decision-making processes within organizations.

Demand

Let $P(y_t,\varepsilon_t)$ be the inverse demand faced by the organization in period t, where y_t is the amount of output produced in the current period and ε_t is a stochastic shock to the demand for the output of the organization. For a given value of ε_t, we assume that this function is twice continuously differentiable and strictly decreasing in y and that the product $P(y,\varepsilon)y$ is strictly concave in y and satisfies the boundary conditions:

$$\lim_{y\to 0}\frac{d}{dy}P(y,\varepsilon)y=\infty, \text{and } \lim_{y\to\infty}\frac{d}{dy}P(y,\varepsilon)y<0.$$

The first condition will ensure that an organization that remains in existence will always want to produce a positive amount of output, and the second condition states that output can effectively be viewed as bounded for any given ε, because the organization will never produce beyond the point where revenues are decreasing in output.

We assume a very simple form for the stochastic process on ε. In particular, we assume that ε_t takes on only one of two values, ε^s or ε^l, where $\varepsilon^s < \varepsilon^l$, with the interpretation that ε^s is the low demand state that will give rise to a small organization, while ε^l is the high demand state that will give rise to a large organization. To generate the standard life-cycle profile of organization size, we assume that when an organization is first created it will have a value of ε equal to ε^s and that over time the state of demand may increase to ε^l. We simplify this process by assuming that the probability that ε increases from ε^s to ε^l is given by the value π^l, which is assumed to be constant over time. We also assume that the process never transits from ε^l back to ε^s. To capture the notion that state ε^l is better than state ε^s, we assume that $P(y,\varepsilon^l) > P(y,\varepsilon^s)$ for all positive values of y and that

$$\frac{d}{dy}P(y,\varepsilon^l)y>\frac{d}{dy}P(y,\varepsilon^s)y$$

for all positive y as well.

If the stochastic evolution of ε just described were a complete description of the uncertainty facing a given organization, then all new organizations would eventually become "large" and remain that way forever. We incorporate the fact that organizations do not last forever by assuming that organizations face an exogenous probability of death and that this probability varies with the state of their demand, ε. In particular, we assume that an organization with demand state ε^i faces a probability λ^i of death. This assumption implies not only that organizations do not last forever, but also that not all new organizations will necessarily become "large."[5] We assume that the realizations of the random variables are independent.

Our assumption about timing is as follows. An organization that produced in period t with demand state ε^i finds out the realizations of the demand and death shocks at the beginning of period t+1 but before any decisions are made in period t+1.

Production

We assume that labor is the only factor of production. The production technology in our model has several key features, which we detail in several steps.

Scale Effects. We assume that there are different ways to organize production and that the optimal way to organize production depends on the scale of production. In general, one could imagine a large set of possible ways to organize production, but given that we are restricting attention to a model with two demand states we also assume that there are only two ways in which production can be organized. We refer to each different way to organize production as a distinct technology. We let $h^i(y)$ denote the labor necessary to produce output y using technology i. Because we will implicitly restrict parameter values so that an organization with demand state ε^s will always use technology 1 and an organization with demand state ε^l will always use technology 2, we will also use s and l to index the two technologies. The idea is that technology 1 is better when producing at a small scale, whereas technology 2 is better when pro-

5 This specification amounts to assuming that there are three demand types, with the third state being an absorbing state in which the organization cannot sell any positive amount of output and still receive a positive price. We identify this third state with death of the organization.

ducing at a large scale. We adopt the following functional forms for the two technologies. We assume that $h^s(y) = a_s y$ for all positive values of y, where $a_s > 0$. The other technology is operable only above some minimum threshold scale, $\bar{y}$, and for $y \geq \bar{y}$ we assume that $h^l(y)$ takes the form $h^l(y) = a_l y$, where $a_l < a_s$. More generally we could specify that this technology can operate at a scale less than $\bar{y}$ but that the average product of labor is sufficiently low that no one would choose to operate it below scale $\bar{y}$. We assume free disposal of output, implying that an organization could choose to operate the large-scale technology but only sell a fraction of the output, though in our analysis we will implicitly assume that this never happens.

We assume that the organization hires labor in a competitive market and hence takes the wage rate as given. In what follows we normalize the wage rate to 1.

Organizational Waste. An important goal for any organization is to use its resources efficiently. The large differences in measured productivity across organizations suggest that organizations vary in the degree to which they accomplish this. Inefficient use of resources may take several forms. We incorporate one particular form of inefficiency, which we refer to as waste. What we have in mind is that in any organization there is potentially some duplication of effort or unnecessary tasks being performed that affect labor productivity in an inframarginal way. In particular, in the context of the technologies described in the previous section, we assume that an organization with waste ϕ has labor requirement $h(y) + \phi$ rather than $h(y)$. The key feature of this waste is that it affects average labor productivity but not marginal labor productivity. One could obviously consider inefficiency that also serves to alter the slope of $h(y)$. Inefficiencies of this form are certainly plausible. We assume inefficiency only of the form as characterized by the parameter ϕ because it is this type of inefficiency that will be central to our analysis. We interpret this inefficiency as reflecting inefficiency in the organizational design and not inefficiency due to workers shirking, for example.

For our purposes there are two key issues associated with these inefficiencies. The first is where they come from, and the second is how organizations can eliminate them. Again, our formulation will be somewhat specialized to isolate a particular effect. In general, one could imagine that inefficiencies stochastically occur within any organization and that it takes organizational resources to get rid of them. One could also assume that organizations devote resources to these activities ex ante to reduce the likelihood that they arise. Our formulation relies on the notion that changes in organizational scale are likely to be associated with the appearance of inefficiencies because the organization is less likely to know how to best use resources as it moves to a new organizational structure.[6] Motivated by this idea, we assume that all organizations operating the small-scale technology do so efficiently, but that whenever an organization switches from the small-scale technology to the large-scale technology it will necessarily move to a positive level of inefficiency that we denote by the parameter $\bar{\phi}$. That is, if an organization used technology s in period $t-1$ and switches to technology l in period t, then their labor requirement function will be $a_l y + \bar{\phi}$ for y $\geq \bar{y}$. We note that it would be straightforward to also assume that a new organization that is operating the small-scale technology for the first time also begins with an inefficiency, but we abstract from this possibility for simplicity. We assume that the level of inefficiency is known to the organization. Although one could consider interesting issues that arise from organizations not having complete information about the state of their efficiency and needing to learn over time about them, we abstract from them here.

Having described how inefficiencies arise, we now turn to the issue of how an organization can get rid of them. We adopt a simple and straightforward formulation. In particular, in any given period an organization makes a discrete decision about whether to try to eliminate inefficiency. Having done so, with probability π^e the organization will decrease its inefficiency to zero in the following period, whereas with probability $1-\pi^e$

[6] Bertschek and Kaiser (2001) present evidence from a data set of German firms that changes in productivity are linked with changes in organizational structure.

the organization will experience no change in its level of inefficiency. Assuming that there is no improvement in efficiency, the organization can continue to try to get rid of the inefficiency in each subsequent period.

Our formulation assumes that a given organization makes stochastic transitions between two levels of efficiency. If we had a large number of organizations all with inefficiency $\bar{\phi}$ and they all continued to try to eliminate this inefficiency until successful, then the average level of inefficiency among this group of organizations would decrease monotonically and approach zero asymptotically. As we will see later in the analysis, this is the pattern that we want to generate. Of course, this pattern could also be generated by having each individual organization experience a monotonically decreasing level of inefficiency rather than the all-or-nothing form that we specify. We have chosen the all-or-nothing form of improvements to simplify the analysis of the model.

In the next subsection we describe in more detail the cost to the organization of trying to reduce its level of inefficiency. Note that we assume that there is no direct cost associated with changing from one technology to another. It would be straightforward to add such a cost but it is not central to the effects that we stress below. Last, our model is related to models of costly adjustment and models of organizational capital, so it is of interest to remark on these relationships. At a general level, the inefficiency associated with change of scale can obviously be interpreted as a form of adjustment cost. We note, however, that our specification differs from most specifications of adjustment costs because the cost does not necessarily disappear in the periods following the adjustment. That is, most models of adjustment costs assume a one-time cost associated with the adjustment, but here the cost is permanent unless the organization takes some actions. The restructuring that takes place within organizations in our model can also be interpreted as a form of investment in organizational capital. However, our model differs from many formulations of organizational capital in that we implicitly assume that this investment in organization capital is a substitute for labor input because it leads to a reduction in labor, whereas most analyses assume that increases in organizational capital lead to increases in the marginal product of labor.

The Role of the Manager. We assume that each organization has one manager and potentially many workers. The labor requirement functions described in the previous subsection should be thought of as specifying the required amount of nonmanagerial labor input. In this regard our model is similar to the standard span-of-control model of Lucas (1978). However, we deviate from that model by allowing a manager to choose between two primary uses of their time. In particular, we assume that managers can devote their time either to facilitating production or to trying to reduce inefficiency. We shall cast this choice as having the manager choose between “producing” or “reorganizing,” which we will denote as $m = p$ and $m = r$, respectively. The cost of having managers devote time to reorganizing is that they are not able to focus on production. We model this cost as a decrease in the efficiency of labor used in the organization. In particular, we assume that if a manager of an organization using large-scale technology with inefficiency $\bar{\phi}$ devotes time to reorganization, then the labor requirement function becomes $(1+\bar{\eta})h^l(y) + \bar{\phi}$, where $\bar{\eta} > 0$ is the efficiency loss associated with the manager not focusing attention on production. The benefit of having the manager focus on reorganizing is that it makes it possible for the organization to be more efficient in the future. As this description makes clear, a key trade-off that a manager faces when making decisions about time allocation is between current efficiency and future expected efficiency. As we will see in the next section, it is this tradeoff and how it interacts with business cycle shocks that is at the heart of our analysis.

We assume a competitive market for (homogeneous) managers. The organization will therefore also take the managerial wage as given, which we also assume to be constant over time. Because an organization cannot function without a manager, managerial compensation is effectively a fixed per-period cost for the organization. We denote this wage by w_m. The only way that an organization can avoid having to hire a manager is to cease to exist. We assume that if an organization chooses this option that it cannot return in the future.

The Organizational Life Cycle

It is straightforward to formulate the optimization problem of the organization just depicted. We do it recursively. The state vector for the organization that remains alive is denoted by $s = (\varepsilon,\phi)$, where ε is the state of demand for its product and ϕ is its level of inefficiency if it chooses to operate the large-scale technology. An organization is always born into the state $(\varepsilon_s,\bar{\phi})$, which is to say that a new organization begins with demand in the low state and an inefficiency level of $\bar{\phi}$. Note that the level of inefficiency begins at $\bar{\phi}$ because, if a new organization were to use the large-scale technology, it would be faced with the inefficiency. But, as noted earlier, as long as it chooses to operate the small-scale technology, it can do so without experiencing any inefficiency. In each period, after observing its current state variable, if the organization remains alive it faces three choices: which technology to use ($i = s$ or l), how much output to produce (y), and how to allocate the manager's time ($m = p$ or r). Given the organization's state vector and choices for each of these decisions, we can determine the current revenues net of payments to nonmanagerial labor that would accrue to the organization, which we will denote by $R(\varepsilon,\phi,i,y,m)$. This function takes the following form:

$$\begin{aligned} (1)\quad & R(\varepsilon,\phi,i,y,m) \\ & = P(y,\varepsilon)y - (1+\bar{\eta}\mathrm{I}_{m=r})h^i(y) - I_{i=l}\phi - w_m, \end{aligned}$$

where $I_{m=r}$ is the indicator function for $m = r$ (i.e., the manager reorganizes) and $I_{i=l}$ is the indicator function for using the large-scale technology.

It is now easy to write the Bellman equation for the maximization problem faced by the organization:

$$(2)\quad V(\varepsilon,\phi) = \max\left\{0, \max_{i,y,m}\begin{bmatrix} R(e,\phi,i,y,m) \\ +\beta(1-\lambda^i)EV(\varepsilon',\phi') \end{bmatrix}\right\},$$

where the outer max reflects the decision of whether to remain active and the inner max reflects the optimal choices, assuming that the organization remains active. We have assumed that if the organization ceases to exist, either through choice in the current period or because of a death shock in the next period, all future returns will be zero. The expectation operator E incorporates two elements. First it incorporates the dynamics in the demand state ε, and second it incorporates the dynamics in the level of inefficiency if the organization chooses to have the manager devote time to reorganization.

Given our assumptions thus far, we cannot rule out some rather extreme or degenerate outcomes that are of little interest. We describe some of these now. In what follows we do not offer any specific conditions on the model specification to rule out these outcomes, but do note intuitively what parameters would be relevant in ruling out certain outcomes.

It is possible that a newly created organization cannot earn positive expected lifetime profits and hence will choose to shut down. In particular, as noted previously, the managerial wage acts like a fixed cost of being in operation, and it is well known that in a model with a fixed cost it is not enough to guarantee positive net revenues from the variable factors, as our earlier assumption on P does. Of course, if a newly created organization is choosing to shut down and there is some cost associated with creating an organization in the first place, then this would imply that new organizations are never created. In an equilibrium context in which consumption is infinitely valued at the margin when consumption is zero, such an outcome could not be an equilibrium outcome. In view of this, it is natural to assume that wages are sufficiently low relative to the price of output that new firms choose to operate. Given our assumptions on the price function, P, it follows that if the expected present discounted value of profits is positive for a newly created organization, then it is positive for any feasible state vector. This property does not necessarily imply that the organization will have positive current-period profits in all states; it is possible that a new organization remains active only because of the possibility of transiting to the higher demand state and that the higher demand state is the only state that is profitable in a static sense. Given our assumptions on P, however, it is true that if an organization remains active it will always choose to produce a positive amount of output, even if current-period profits are negative.

The model has been constructed to focus on the change in scale of production and the associated change in organization structure that occurs as organizations successfully mature. Given a price function, P, however, if the value of $\overline{y}$ is sufficiently large then no organization will ever choose to operate the large-scale technology; if $\overline{y}$ is too small, then all organizations will choose to operate the large-scale technology. In the context of our model, neither of these cases is particularly interesting. So, in what follows, we assume that it is optimal for an organization in demand state ε^s to operate the small-scale technology and for an organization in the demand state ε^l to operate the large-scale technology. We note, as a feature of our specification, that it is not possible to eliminate future inefficiency while currently operating the small-scale technology. It follows that, even if an organization decides to reorganize and thereby experiences the current loss of efficiency associated with η, the organization will still choose to operate the large-scale technology.

Last, it is also possible that the values of ϕ and $\overline{\eta}$ are such that no organization would ever choose to reorganize. This could happen if the level of inefficiency (i.e., ϕ) is sufficiently small relative to the foregone productivity (i.e., $\overline{\eta}$ is large) or the probability of failure in reorganization (i.e., $1-\pi^e$). Conversely, if the size of the inefficiency is sufficiently large relative to the cost of eliminating it, then an organization would always choose to reorganize. Because the case of no reorganization is not very interesting in the context of our model, we assume in what follows that we are in a region of parameter space in which organizations do sometimes choose to reorganize.

Conditional on assuming that a newly created organization chooses to remain in existence, that organizations in the low (high) demand state operate the small- (large-) scale technology, and that organizations sometimes choose to eliminate inefficiency, it is fairly easy to characterize the life cycle dynamics that emerge. In particular, any newly created organization will operate the small-scale technology and hire an amount of labor that we denote by h^s, producing output denoted by y^s. Over time there are three things that may happen to this organization. It may receive a shock and cease to exist, it may remain in the same position and hence continue to hire h^s workers, or it may experience a shock that increases its demand to the high state.

If it is sometimes optimal to try to eliminate inefficiency, then because the organization's problem is recursive, it must be optimal to do it the first time the organization reaches the high demand state. While in the high state and reorganizing, the organization is employing the large-scale technology and hires labor that we denote by h^r, producing output denoted by y^r. Even though the manager's devotion of time to reorganizing lowers the marginal product of labor, it still must be the case that $h^r > h^s$. To see this, note the following. First, this organization must be producing at least $\overline{y}$ units of output, which must exceed the amount of output produced in the low demand state; otherwise it would not have been optimal to use the small-scale technology in the low demand state. Now, if it was possible to produce more output with less labor, then the organization could have chosen this combination in the previous state and chosen not to sell all of the output produced. It follows that h^r must exceed h^s. It follows that if an organization experiences an improvement in its demand state, it increases both its labor input and its output. Note that we cannot say anything about what happens to average labor productivity. Productivity will depend on the magnitudes of the parameters $\overline{\eta}$ and $\overline{\phi}$.

An organization in the high demand state that is reorganizing can in turn experience three different transitions. First, it may receive a bad shock and cease to exist. Second, it may be unsuccessful in eliminating inefficiency and remain in the same state, in which case it chooses the same actions again. Third, it may be successful in eliminating the inefficiency. We denote the levels of y and h that result in this case as h^l and y^l. How do the values of h^l,y^l and y^l/h^l compare with the corresponding values from earlier in the life cycle? The first-order condition for current-period choice of output combined with our assumptions on P implies that output will definitely increase when the manager focuses on production rather than reorganization. This occurs because the term $(1 + \bar{\eta})$ goes away from the first-order condition. However, it is ambiguous whether this leads to an increase or decrease in h, for two reasons. First, even if the only effect were the improved efficiency associated with the managerial time allocation, the effect on labor, as opposed to output, will depend on the elasticity of the demand function. Second, the elimination of the inefficiency measured by $\bar{\phi}$ necessarily implies a decrease in labor in the amount of $\bar{\phi}$. However, although the effect on h is ambiguous, it is easy to see that independently of what happens to h, average labor productivity will necessarily increase.

Although with our implicit assumptions on parameter values that the organization will never choose to postpone reorganization once it reaches the high demand state, we can still ask what levels of h and y would be optimal if it chose to do so. Denote these levels by h^p and y^p. We ask how h^p and h^l compare. In making this comparison we are assuming that the current-period marginal efficiencies are the same because in both cases the manager is focusing on production. Given our formulation, it follows that output will be the same in each case (i.e., $y^p = y^l$). However, because of the waste in the former case, we know that $h^p > h^l$, which will be of particular interest later because it states that an inefficient firm that chooses to postpone reorganization will necessarily shed workers in the future.

As a final remark in this section, we note that our model emphasizes the restructuring that accompanies growth of an organization. It seems equally plausible that restructuring within shrinking organizations would also be of importance. For the implications that we stress, we believe that similar results would emerge from this situation as well, so we have chosen to focus on growing organizations purely for simplicity.

THE MODEL WITH TEMPORARY SHOCKS

The previous model considered the decision-making of an organization over its life cycle. The only shocks in that model were highly persistent, and we interpreted them to be organization specific. In this section we add purely temporary shocks to the model and examine how these temporary shocks influence the life cycle dynamics that we studied earlier. Although our model is purely decision theoretic, we will interpret the i.i.d. shocks that we introduce as reflecting (aggregate) business cycle shocks.

Formally, we assume a second shock that influences demand; for simplicity we assume that this shock is multiplicative, so that we write the inverse demand function facing the organization as $\varepsilon_{2t}P(y_t,\varepsilon_{1t})$, where P is the same function as described earlier, ε_{1t} is the shock that we previously labeled as ε_t, and ε_{2t} is an i.i.d. shock that is drawn from a distribution with cdf $F(\varepsilon_2)$. We assume that realizations of ε_2 lie in the interval $[\varepsilon_{\min},\varepsilon_{\max}]$ and that this interval contains 1 in its interior. Note that if $\varepsilon_{2t} = 1$ for all t, then the model is identical to that considered previously. All other aspects of the environment are left unchanged.

Proceeding as before, we define the revenue associated with decisions in a particular state as $R(\varepsilon_1,\varepsilon_2,\phi,i,y,m)$. This function takes the following form:

$$\begin{aligned} &R(\varepsilon_1,\varepsilon_2,\phi,i,y,m) \\ &= \varepsilon_2 P(y,\varepsilon_1)y - (1+\bar{\eta}I_{m=r})h^i(y) - I_{i=l}\phi - w_m, \end{aligned} \tag{3}$$

where $I_{m=r}$ is the indicator function for $m = r$ (i.e., the manager reorganizes) and $I_{i=l}$ is the indicator function for using the large-scale technology. It is immediate that both R and R_y are increasing in ε_2.

It is again easy to write the Bellman equation for the maximization problem faced by the organization:

$$V(\varepsilon_1,\varepsilon_2,\phi) = \max\left\{0, \max_{i,y,m}\begin{bmatrix} R(\varepsilon_1,\varepsilon_2,\phi,i,y,m) \\ +\beta(1-\lambda^i)EV(\varepsilon_1',\varepsilon_2',\phi') \end{bmatrix}\right\}. \tag{4}$$

For future reference it is worthwhile to elaborate on the expected value term in more detail. As noted earlier, this expectation takes into account the evolution of the exogenous shocks as well as the evolution of the inefficiency variable in response to the decision about reorganization. If $\varepsilon_1 = \varepsilon^s$, then the only value of ϕ of interest is $\phi = \bar{\phi}$.

Assuming this configuration plus an arbitrary value for ε_2, the term for the next period's value in the Bellman equation becomes

$$\beta(1-\lambda^i)EV(\varepsilon_1',\varepsilon_2',\phi') \tag{5}$$
$$= \beta(1-\lambda^s)\left[\begin{array}{l}\pi^l \int V(\varepsilon^l,\varepsilon,\bar{\phi})dF(\varepsilon) \\ \qquad +(1-\pi^l)\int V(\varepsilon^s,\varepsilon,\bar{\phi})dF(\varepsilon)\end{array}\right].$$

If the organization has $\varepsilon_1 = \varepsilon^l$, then the only efficiency value of interest is still $\phi = \bar{\phi}$, because otherwise there is no need for a decision about reorganization and the problem becomes static. For an arbitrary value of ε_2, if the organization decides to reorganize, then the future term in the Bellman equation becomes

$$\beta(1-\lambda^i)EV(\varepsilon_1',\varepsilon_2',\phi') \tag{6}$$
$$= \beta(1-\lambda^l)\left[\begin{array}{l}\pi^e \int V(\varepsilon^l,\varepsilon,0)dF(\varepsilon) \\ \qquad +(1-\pi^e)\int V(\varepsilon^l,\varepsilon,\bar{\phi})dF(\varepsilon)\end{array}\right];$$

whereas, if it chooses not to reorganize, then the same term becomes

$$\beta(1-\lambda^i)EV(\varepsilon_1',\varepsilon_2',\phi') = \beta(1-\lambda^l)\int V(\varepsilon^l,\varepsilon,\bar{\phi})dF(\varepsilon). \tag{7}$$

As was true in the previous subsection, depending on parameter values there are various forms that the optimal decision rules may take. We modify our previous assumptions marginally, so we now assume that when $\varepsilon_1 = \varepsilon^s$ the organization will choose to remain active independently of the value of ε_2.[7] We furthermore assume that when an organization experiences an increase in its demand state from ε^s to ε^l that there is at least some interior value of ε_2 for which the organization would choose to reorganize.

Finally, we also place an implicit assumption on the size of the shocks to ε_1 and ε_2. In particular we assume that life-cycle shocks are much larger than business-cycle shocks. The significance of this is that we assume that when $\varepsilon_1 = \varepsilon^s$ the organization does not wish to operate the large-scale technology independently of the realization of ε_2. Similarly, we assume that when $\varepsilon_1 = \varepsilon^l$ the organization never chooses to operate the small-scale technology independently of the realization of ε_2. This model is identical to the model of the previous section if we assume that $\varepsilon_{\min} = \varepsilon_{\max} = 1$. It follows that, if the previous model implies technology i is operated only in demand state i, this model will also generate this result if the range of ε_2s is not too large.

We are now able to prove our main result, which is that the decision to reorganize when in state $(\varepsilon^l,\varepsilon_2,\bar{\phi})$ is characterized by a reservation value of ε_2, with the property that it is optimal to reorganize if $\varepsilon_2 < \varepsilon_2^*$, and not to reorganize if $\varepsilon_2 > \varepsilon_2^*$. The intuition for the result is simple: It basically reflects intertemporal substitution of reorganization. Reorganization imposes a cost today in terms of foregone efficiency of labor, but offers a future gain in reducing waste. If ε_2 is i.i.d., then future gains are the same independently of the current value of ε_2. But, we will show that the current-period cost of reorganizing is increasing in the amount of production desired in the event of not reorganizing, which in turn is increasing in ε_2.

We now establish these results. Consider an organization in state $(\varepsilon^l,\varepsilon_2,\bar{\phi})$. Let y^p denote the optimal level of production if the organization were to choose not to reorganize, and let y^r denote the optimal level of production were the organization to choose to reorganize. Conditional on deciding whether to reorganize, note that the resulting decision about the optimal choice of y is static and can be represented as

$$W(\varepsilon_2,\eta) = \max_y \left\{\varepsilon_2 P(y,\varepsilon^l)y - (1+\eta)h(y)\right\}, \tag{8}$$

where η takes on the value 0 in the event of $m = p$ and $\eta = \bar{\eta}$ in the event that $m = r$. We denote the optimal choice of y as $y(\varepsilon_2,\eta)$. Note that the first-order condition that defines this function is

$$\varepsilon_2\left[yP_1(y,\varepsilon^l) + P(y,\varepsilon^l)\right] = (1+\eta)h'(y). \tag{9}$$

Given our assumptions, it follows trivially that the optimal value of y is increasing in ε_2 and decreasing in η.

Let $V^p(\varepsilon^l,\varepsilon_2,\bar{\phi})$ and $V^r(\varepsilon^l,\varepsilon_2,\bar{\phi})$ be the resulting optimal values obtained from choosing not to

---

[7] More generally, it would be straightforward to allow for the possibility that the organization will not remain in operation for all realizations of ε_2. In this case there will be a reservation value of ε_2 that dictates whether the organization remains.

reorganize and to reorganize, respectively, assuming that output is chosen optimally in each case. Now consider the difference between these two values. Using V^p and V^r to denote these functions, direct substitution gives

$$V^p\left(\varepsilon^l,\varepsilon_2,\bar{\phi}\right)-V^r\left(\varepsilon^l,\varepsilon_2,\bar{\phi}\right)=R^p-R^r+\beta\left(1-\lambda^l\right)\left\{\begin{array}{l}\int V\left(\varepsilon^l,\varepsilon_2,\bar{\phi}\right)dF(\varepsilon)\\ -\left[\begin{array}{l}\pi^e\int V\left(\varepsilon^l,\varepsilon_2,0\right)dF(\varepsilon)\\ \quad+\left(1-\pi^e\right)\int V\left(\varepsilon^l,\varepsilon_2,\bar{\phi}\right)dF(\varepsilon)\end{array}\right]\end{array}\right\}, \tag{10}$$

where we have written $R^p = R(\varepsilon^l,\varepsilon_2,\bar{\phi},l,y^p,p)$ and $R^r = R(\varepsilon^l,\varepsilon_2,\bar{\phi},l,y^r,r)$. Note that the value of the terms involving integrals are all independent of ε_2. Denote these terms by the constant A. Moreover, the difference in the two revenues can be reduced to

$$R^p-R^r=W\left(\varepsilon_2,0\right)-W\left(\varepsilon_2,\bar{\eta}\right). \tag{11}$$

It follows that equation (10) can be written as

$$\begin{aligned}&V^p\left(\varepsilon^l,\varepsilon_2,\bar{\phi}\right)-V^r\left(\varepsilon^l,\varepsilon_2,\bar{\phi}\right)\\&=\left[W\left(\varepsilon_2,0\right)-W\left(\varepsilon_2,\bar{\eta}\right)\right]+A.\end{aligned} \tag{12}$$

This equation is intuitive. The term in square brackets is the current-period cost of reorganizing: It represents the loss in current revenue associated with having the manager devote time to reorganizing. The term A is the future benefit to reorganizing. Given our assumption, this is simply a positive number that is independent of ε_2.

We can now easily show that this difference is increasing in ε_2. It is sufficient to show that the term in square brackets is increasing in ε_2. Differentiation gives

$$\frac{d}{d\varepsilon}\left[W\left(\varepsilon_2,0\right)-W\left(\varepsilon_2,\bar{\eta}\right)\right]=W_1\left(\varepsilon_2,0\right)-W_1\left(\varepsilon_2,\bar{\eta}\right), \tag{13}$$

so it is sufficient to show that $W_{12} < 0$. By definition,

$$\begin{aligned}&W\left(\varepsilon_2,\eta\right)\\&=\varepsilon_2 P\left(y\left(\varepsilon_2,\eta\right)\right)y\left(\varepsilon_2,\eta\right)-\left(1+\eta\right)h\left(y\left(\varepsilon_2,\eta\right)\right).\end{aligned} \tag{14}$$

Using the envelope condition, we have that

$$W_2\left(\varepsilon_2,\eta\right)=-h\left(y\left(\varepsilon_2,\eta\right)\right)<0. \tag{15}$$

It then follows that

$$W_{12}\left(\varepsilon_2,\eta\right)=-h'\left(y\left(\varepsilon_2,\eta\right)\right)y_1\left(\varepsilon_2,\eta\right). \tag{16}$$

Since h is increasing and the solution for y is increasing in ε_2, it follows that $W_{12} < 0$ and hence $V^p(\varepsilon^l,\varepsilon_2,\bar{\phi}) - V^r(\varepsilon^l,\varepsilon_2,\bar{\phi})$ is increasing in ε_2. If the benefit to not reorganizing is monotonic in ε_2, it follows that the optimal reorganization strategy is to employ a reservation value, as previously stated.

If the reservation value is equal to $\varepsilon_{\max}$ (i.e., the upper support of the distribution of the temporary shocks), then the organizational dynamics in this model are qualitatively the same as in the previous subsection. That is, whenever a small organization gets an improvement in their idiosyncratic demand state ε_1, they immediately choose to reorganize and continue to do so until the reorganization is successful. Fluctuations in ε_2 will lead to additional fluctuations in their labor input and output, but the life-cycle dynamics will be similar.

However, if the reservation value ε_2^* is interior to the interval $[\varepsilon_{\min},\varepsilon_{\max}]$, then qualitatively different dynamics can emerge. In this scenario, if an organization in the small idiosyncratic demand state receives a shock that raises it to the large idiosyncratic demand state, the organization may or may not decide to reorganize at that point. In particular, if ε_2 is sufficiently high, then the organization will choose to postpone the decision to reorganize to take advantage of the current temporarily high demand. As stated earlier, the organization will engage in intertemporal substitution of managerial actions. And, following from our discussion of the previous model, we know that an organization that chooses to postpone reorganization will necessarily employ less labor in the future when it does successfully reorganize, even holding the value of ε_2 constant.

Extension to the Case of Persistent Shocks

The previous analysis has assumed that the shock ε_2 is i.i.d. over time. Of course, if one wants to think of the ε_2 shock as proxying for business cycle movements in the demand faced by an individual organization, then the i.i.d. assumption is not very appealing. A well-documented property of business cycles is that they are persistent, in the sense that if the economy is above trend today then we also expect it to be above trend next period. In view of this, it is of interest to ask whether our result about the reservation value will extend to the case of persistent shocks. In fact, the argument is easily extended.

In the i.i.d. case, we argued that the current cost of reorganizing is increasing in the current value of ε_2 and that the expected future benefit of reorganizing is independent of the current-period value of ε_2. The first statement is independent of whether the realizations of ε_2 are i.i.d. or not. However, a key observation about the structure of our model is that the benefit to successful reorganization is in fact independent of future realizations of ε_2. The reason for this is that, in our model, successful reorganization does not influence the marginal product of labor, and as a result an efficient organization and an inefficient organization will choose the same level of output conditional on having the same managerial time allocation. The only effect on profit is from saving labor in the amount of $\overline{\phi}$ for each future period that the organization remains in existence, and this saving is independent of all future realizations of ε_2.

The additional issue that needs to be addressed in the context of a model with persistent shocks to ε_2 is the following: An organization faced with a current realization of ε_2 can also consider the possibility of waiting a period to reorganize. If the benefit from waiting increases as ε_2 decreases, then the reservation property might not be preserved. In the i.i.d. case, the benefit from waiting is actually increasing in ε_2, because next period's expected one-period cost of reorganizing is independent of ε_2; thus, a higher current value of ε_2 indicates lower expected costs in the future, whereas a low value of ε_2 indicates higher expected costs in the future.

With this in mind we can present the alternative characterization of the decision to reorganize. In particular, let $C(\varepsilon) = W(\varepsilon,0) - W(\varepsilon,\overline{\eta})$ be the gain this period from not reorganizing. Let $G = \beta\pi^e\overline{\phi}/(1-\beta(1-\lambda^l))$ be the (expected) gain from choosing to reorganize today. Letting $H(\varepsilon)$ be the benefit of today's managerial choice relative to reorganizing today, we have that $H(\varepsilon)$ satisfies

$$(17)\quad H(\varepsilon) = \max\left\{C(\varepsilon) + \beta\int H(\varepsilon')F(d\varepsilon',\varepsilon),0\right\},$$

where the first term indicates the gain from not reorganizing today and the second term indicates that if the manager chooses to reorganize then the gain is clearly zero. We know that $C(\varepsilon)$ is increasing in ε from our previous analysis. If we knew that the term $C(\varepsilon) + \beta\int H(\varepsilon')F(d\varepsilon',\varepsilon)$ was increasing in ε, the current realization of the shock, then we could easily conclude that the reservation property holds. As stated earlier, note that if ε is i.i.d., then the integral is independent of ε and the property holds based on the property of C. However, finding conditions under which the integral is increasing in ε is a standard problem in dynamic stochastic models. In particular, if we assume that

$$(18)\qquad \int g(\varepsilon')F(d\varepsilon',\varepsilon)$$

is increasing in ε for any increasing function, then we can easily show that the value function H will in fact be increasing and hence that the integral has the desired property.

Loosely speaking, a process for ε_2 that implied mean reversion would tend to satisfy this property. This is because a high value of ε_2 today implies that future values of ε_2 will be lower, implying that it is beneficial to wait to reorganize. On the other hand, a very low value of ε_2 today implies that future values of ε_2 will be higher, implying that there is greater incentive to reorganize today rather than in the future.

We conclude that the reservation value property for the optimal reorganization decision will also hold in the case of persistent shocks under reasonable conditions.

IMPLICATIONS FOR BUSINESS CYCLE DYNAMICS

Our formal analysis has considered only the decision problem of an individual organization that takes demand for its product as given, assuming that wages for workers and managers and the real interest rate are constant over time. Such a model can be cast in an industry equilibrium setting, as in Hopenhayn and Rogerson (1993), in which the organizational dynamics that we describe can capture the steady-state dynamics of a general equilibrium model in which all shocks are idiosyncratic (i.e., there are no aggregate shocks). If one introduces aggregate shocks into such a model, then one would need to take into account the effect that these shocks have on wage rates and the real interest rate. Veracierto (2002) and Thomas (2002) are examples of models in which these general equilibrium effects are considered. Hence, in its current form the model is really not appropriate to discuss how the economy responds to aggregate shocks. Nonetheless, in this section we want to discuss some potential effects suggested by previous analysis for business cycle dynamics. We leave development of the appropriate framework and the associated formal analysis for future work.

Consider the following situation. We have a unit mass of entry of new organizations each period, each of which enters into the individual state $(\varepsilon^s, \bar{\phi})$. We consider the ε_2 shock to be common to all organizations, while the ε_1 shock is idiosyncratic, and trace out the evolution of the economy assuming that wage rates and the interest rate remain constant. The first observation that we want to stress is that the aggregate state of this economy will be the realization of the aggregate shock ε_2 and the distribution of organizations across individual states. Entering a given period there are three types of organizations: those that are in the state $(\varepsilon^s, \bar{\phi})$ (which we call type 1), those that are in the state $(\varepsilon^l, \bar{\phi})$ (which we call type 2), and those that are in the state $(\varepsilon^l, 0)$ (which we call type 3). We let μ_i be the mass of firms in each of the three states. The evolution of the μ_is is affected by the realization of the aggregate shock because it determines whether organizations in state 2 will be reorganizing, thereby influencing the probability that an organization transits to state 3. In particular, let μ_{it} be the distribution of active organizations in period t. Then, if $\varepsilon_{2t} > \varepsilon_2^*$, we have the following:

$$
\begin{aligned}
\mu_{1t+1} &= \left(1-\lambda^s-\pi^l\right)\mu_{1t}+1 \\
\mu_{2t+1} &= \left(1-\lambda^l\right)\mu_{2t}+\pi^l\mu_{1t} \\
\mu_{3t+1} &= \left(1-\lambda^l\right)\mu_{3t}.
\end{aligned} \tag{19}
$$

If, on the other hand, we have $\varepsilon_{2t} < \varepsilon_2^*$, then the distribution will evolve according to

$$
\begin{aligned}
\mu_{1t+1} &= \left(1-\lambda^s-\pi^l\right)\mu_{1t}+1 \\
\mu_{2t+1} &= \left(1-\lambda^l-\pi^e\right)\mu_{2t}+\pi^l\mu_{1t} \\
\mu_{3t+1} &= \left(1-\lambda^l\right)\mu_{3t}+\pi^e\mu_{2t}.
\end{aligned} \tag{20}
$$

Two simple conclusions can be drawn from these laws of motion. First, note that the law of motion for μ_{1t} is independent of the realization of ε_2. It follows that, if entry is constant as we have assumed, the value of μ_1 will approach a constant and will not be affected by realizations of ε_2. The constant fraction of type 1 organizations is easily computed to be $\lambda^l/(\lambda^l+\pi^l)$.The second point to note is that the remaining mass of organizations will be split between type 2 and type 3 organizations and that this division will depend on the history of the ε_2 realizations. Specifically, if ε_2 remains above ε_2^*, then there is a greater buildup of type 2 organizations at the expense of type 3 organizations. It follows that the longer the aggregate shock remains above the reservation value, the greater will be the buildup of type 2 organizations. In what follows, we illustrate the potential effects that this can have on how the economy responds to subsequent shocks.

A Reduced-Form Example

For present purposes, the most effective way to illustrate the interaction between the distribution of organizations and the response of the economy to a given sequence of aggregate shocks is with a very specific reduced-form example. Consider an economy at time 0 with unit mass of organizations that are distributed across types. We assume that the stochastic process for ε_2 has a sequence of realizations of the following form. In period 0 the economy is hit by a (very) negative value of ε_2 but, subsequent to this, experiences a constant and gradual increase back toward its unconditional mean value. We assume that it takes the economy 20 periods to reach this value, at which time it remains there. This type of realization could be thought of as tracing out the impulse response function in the presence of a mean-reverting process.

We assume the following reduced-form properties in our example. First, we assume that reorganization is optimal for this entire range of realized ε_2 values. Second, we assume that when an organization reorganizes successfully, the effect on labor input is a decrease of e^f, holding ε_2 constant, and is independent of the level of ε_2. Third, we assume that, for each period in which ε_2 is increasing, the aggregate effect of this on labor input is an increase in labor input of e^h, which we assume is distributed across organizations according to size.[8] We assume that in period 0 a small organization employs 10 workers and that a large reorganized organization employs 100 workers. By assumption, a large organization in the process of reorganizing will employ $100 + e^f$ workers. The probabilities π^l,π^e,λ^s, and λ^l are as before.

Our goal here is to illustrate the potential for the initial distribution of organizations to influence the resulting response of employment to a given sequence of shocks. With this in mind, we simulate the implied path of aggregate employment for several different initial conditions. As already noted above, with a constant rate of entry of new organizations, asymptotically there will be a constant mass of organizations and a constant fraction of them will be of type 1 (i.e., in the low demand state with inefficiency $\overline{\phi}$). In view of this we always consider the initial fraction of type 1 organizations to correspond to this fraction. We normalize the total mass to equal 1 and hence assume that $\mu_{10} = \lambda^l/(\lambda^l + \pi^l)$. As noted earlier, however, at any given point the distribution of the remaining organizations between type 2 and type 3 will be influenced by the previous history of realizations of ε_2, and we therefore consider different scenarios for how this remaining mass is allocated across type 2 and type 3 organizations.

For the paths shown in Figure 1 we have set $e^h = 0.1$ and $e^f = 10$. Setting $e^h = 0.1$ amounts to assuming that the accumulated increases in ε_2 over the 20 periods would increase aggregate employment by roughly 4 percent over the course of 20 periods, holding all else constant. We set $\varepsilon^f = 10$, implying that successful reorganization leads to a reduction in employment of roughly 10 percent. We set $\lambda^l = \pi^l = 0.0025$, implying that over the course of the 20 periods the accumulated probability of failure for a large organization or success for a small organization is roughly 5 percent. Finally, we set $\pi^e = 0.05$, which implies an expected duration of 20 periods for successful reorganization. Although we have offered these quantitative guides to thinking about the parameter selections, we also emphasize that this example is purely illustrative. We leave a rigorous quantitative assessment of the economic mechanisms described here for future work.

Figure 1 shows the employment paths that result for three different initial values of μ_2. In each case, the curve represents deviations from means for each path to focus attention on the implications for timing. The figure shows that as μ_{20} increases it takes longer to reach the turning point of employment. This result is intuitive. The greater the value of μ_{20}, the greater is the total amount of reorganization that needs to be done. As this reorganization takes place, successful organizations will be shedding labor. This labor shedding is an opposing force to the increases in labor associated with the gradual improvement in the aggregate shock ε_2. A key point to note is that in our model, reorganization is potentially a long-lasting process in the sense that in any given period only a given fraction π^e of the remaining reorganization will be carried out. In fact, it is interesting to note the implications of the extreme case in which $\pi^e = 1$. In this case all of the "accumulated" reorganization will be carried out in the first period and there will be a large drop in employment, but after this we will see a continual increase in aggregate employment as ε_2 increases. Hence, a large amount of accumulated reorganization will simply lead to a very large one-time drop in employment, but will not lead to a delayed turning point for aggregate employment.

To understand the dynamics of the opposing forces, Figure 2 shows the time paths of firing and

[8] One could interpret this reduced form as reflecting a log linearization of the individual demand for labor functions.

Figure 1

Effect of Initial Conditions on Employment

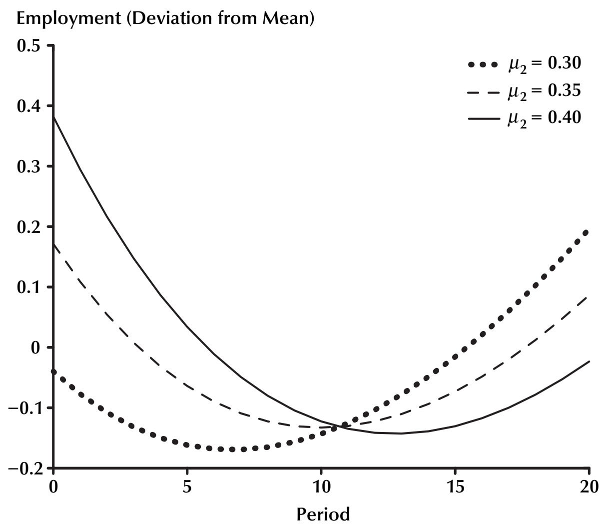
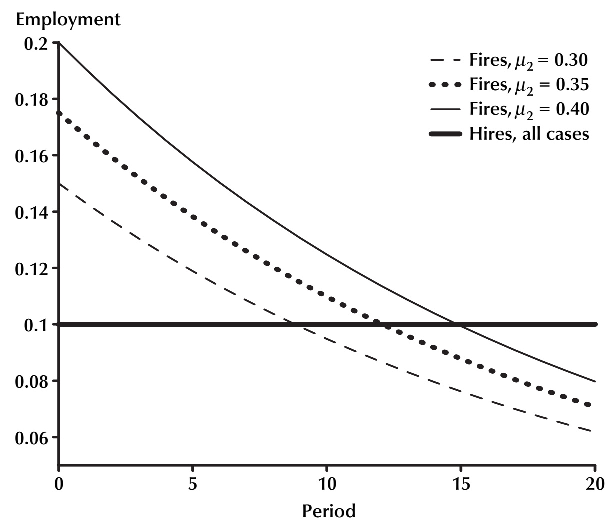


Figure 2

Initial Conditions and Hiring and Firing



hiring for each of the three scenarios considered in Figure 1.[9]

In all three cases the hiring associated with the improvement in ε_2 is constant over time and equal to 0.1. However, although each economy has the same fundamentals, the variation of the initial distribution of μ implies that the time path of fires associated with successful reorganization will be different. As Figure 2 shows, the curves are effectively parallel shifts of each other, with the highest value of μ_{20} associated with the highest level of firing. Because successful reorganization takes time, the path of firing is fairly drawn out. The important feature to note is that aggregate employment will continue to drop as long as the firing curve lies above the hiring curve. And because a higher value of μ_{20} raises the firing curve but leaves the hiring curve unchanged, it illustrates how restructuring can influence the point at which aggregate employment begins to increase.

We should emphasize that the dynamics that we have just traced out are obviously not definitive predictions of the model. As noted earlier, whether reorganization leads to labor shedding depends on parameter values. The main point we want to emphasize is that the model suggests a mechanism that can produce these types of dynamics.

Firing versus Hiring

The results of the numerical example reported in the previous section show that aggregate employment growth is slower because many organizations are reducing the size of their workforces. This finding suggests that slow aggregate employment growth should be associated with high separation rates. In fact, Shimer (2005a,b) argues that the slow aggregate employment growth in the recent recovery is due to a low rate of hiring and not to a high rate of separations, including layoffs. In this subsection we discuss how our mechanism could be made consistent with this observation. First note that the curves that we labeled as hires and fires do not actually correspond to their counterparts in the data. Rather, these two curves simply reflect two different forces, one leading to lower employment and one leading to greater employment. If the force

[9] An organization that successfully reorganizes may, of course, choose to fire fewer than 10 workers and not hire any new workers; so, when measured, hiring and firing in the economy may not reproduce these curves.

leading to lower employment called for a decrease of ten workers in an organization and the force leading to higher employment called for an increase of eight workers in that same organization, then we would expect this to show up as a situation with zero hires and two separations.

We argue here that a simple extension of our model can potentially explain why it is low hiring rather than high separations that seems to be the proximate cause of the slow employment growth. If an organization finds itself with too many workers, but expects that over time these workers will be needed, then if the workers possess some valuable organization capital it may be optimal for the firm to simply keep the workers around and let any decreases in employment occur through attrition. This policy could produce a pattern of relatively stable separation rates coupled with an extended period of very low hiring rates. This possibility would become even more relevant if the organizations that grew during the preceding expansion are also the ones that are expected to grow the most in the near future, because they would then represent both the organizations with the most unneeded labor currently and the ones most expected to expand their labor forces in the near future.

General Equilibrium Considerations

Although our analysis has considered the decision problems of only individual organizations and then aggregated holding wage and interest rates constant, it is worthwhile to discuss how general equilibrium considerations could possibly affect the types of outcomes that we are emphasizing. In particular, it is important to emphasize this issue in the context of the literature related to the work of Caballero and Engel (1999) referred to in the introduction. In that paper, they argue that changes in the distribution of individual firm state variables is important in influencing how the economy responds to shocks. However, the work of Veracierto (2002) and Thomas (2002) show that, in general equilibrium versions of the model, the effects of interest rate movements basically offset the partial equilibrium effects.

There are two main issues that arise. The first concerns how changes in prices might impact the incentives for intertemporal substitution in our decision problem, as represented in our key analytic result showing the existence of a reservation value of ε_2 for the reorganization decision. With constant wages and interest rates, reorganization will be shifted away from periods of high economic activity and toward periods of low economic activity. If real wages are procyclical, they could generate an opposing force to our intertemporal substitution. If wages are higher in periods of high economic activity and shocks are persistent, then this produces a cost of not reorganizing that is procyclical. Simply put, the benefit of shedding labor is greater if wages are higher.

General equilibrium effects could also operate through changes in the real interest rate. However, if one views the case of procyclical real interest rates as the case of primary interest, then this effect will actually reinforce our result. If current real interest rates are high, then current-period costs are amplified and future benefits are attenuated, increasing the incentive to postpone reorganization in good times.

One could plausibly argue that some other margins that we have assumed to be constant over time would also exhibit variability over the cycle. For example, although we assumed that transition probabilities are constant over time, one could argue that the probability of a small-scale organization becoming successful increases in good times (i.e., that π^l is higher in good times). This aspect by itself would tend to accentuate the effects that we have emphasized, because this will lead to a larger buildup of type 2 organizations during good times. Similarly, if entry is higher in expansions, then this aspect will also tend to accentuate the buildup of type 2 organizations.

A second issue that must be addressed in a more complete model is why the workers that are being released due to restructuring do not find employment somewhere. One possible channel is standard intertemporal substitution effects. When an organization restructures, the shift of managerial time away from production and toward restructuring leads to a decrease in the marginal product of labor. A second channel that is not in our model but could be important is that the organizations with the greatest expected increases in employment in the future may be those that have recently experienced large increases. If this is true, it could be that the organizations that decide to restructure are the same organizations that will eventually add the most workers. One could imagine a more detailed model in which an organization does not add to its existing labor force at the same time that it is trying to reorganize. Hence, the decision to reorganize is implicitly a decision to postpone new hires. Last, if one were to imbed our model into a model in which it takes time for workers to move from one organization to another, then a long-lasting increase in separations would also lead to a long-lasting decline in employment.

A CLOSER LOOK AT JOBLESS RECOVERIES

In this section, we argue that the insights derived from the preceding discussion may be relevant for understanding some features of business cycle dynamics and that in particular they may be very relevant for the discussion of the phenomenon that has become known as the jobless recovery. As noted in the introduction, many individuals have coined the term "jobless recovery" to describe the apparent slow growth in employment following the troughs of the two most recent recessions, in 1991 and 2001. Viewed in a broader perspective, the obvious implication of such a description is that not all business cycles are alike, which is an old and recurring theme in the business cycle literature. Burns and Mitchell (1946) were among the first to systematically measure the business cycle and argued that business cycles bear a remarkable similarity to each other along many dimensions. In particular, they developed the notion of a reference cycle to represent the "typical" business cycle. Influenced by this work, Lucas (1977) argued that a key stylized fact is that all business cycles are the same from the perspective of qualitative comovement of series. At the same time, there are many instances in which researchers have argued that some particular business cycle exhibits properties that distinguish it from its predecessors, while others have argued that the business cycle phenomenon is slowly changing over time.[10]

The discussion of the previous section suggested that after the end of a long expansion, employment may take longer to start to increase once again. This argument is consistent with the fact that each of the past two recessions has exhibited a relatively long period before employment began to increase, because each of the past two expansions has been extremely long by historical standards. However, there is another episode in the postwar period that would seem to be relevant and that is the recession of 1969-70, which also followed a very long expansion. If the channel that we point to is quantitatively important, then this period should also have produced a "jobless recovery." The goal of this section is to argue that the evidence is indeed consistent with this prediction. In particular, we will argue that there are three recessions in the postwar period that stand out as distinct from the others in terms of the dynamics for employment in the subsequent recovery: 1969, 1991, and 2001. The material presented here draws on the results presented in Koenders (2005), which provides a much more thorough analysis.

A Review of Schreft and Singh

It is useful to begin with a summary of the analysis of Schreft and Singh (2003). They carry out the following calculation: They start with seasonally adjusted data for employment from the Bureau of Labor Statistics Establishment Survey and then identify the level of employment at each of the National Bureau of Economic Research (NBER) turning points that corresponds to the end of a recession. For each recovery, they plot the percentage change in employment from the turning point that occurs over the subsequent 12 months. Figure 3 is equivalent to the figure that they produce except that we have included the two recessions from the 1950s in our analysis and we have time-aggregated the employment data to quarterly frequency.

But our Figure 1 tells the same story as Chart 1

[10] A related but distinct issue is the extent to which business cycles have become less frequent.

Figure 3

The Schreft-Singh Finding

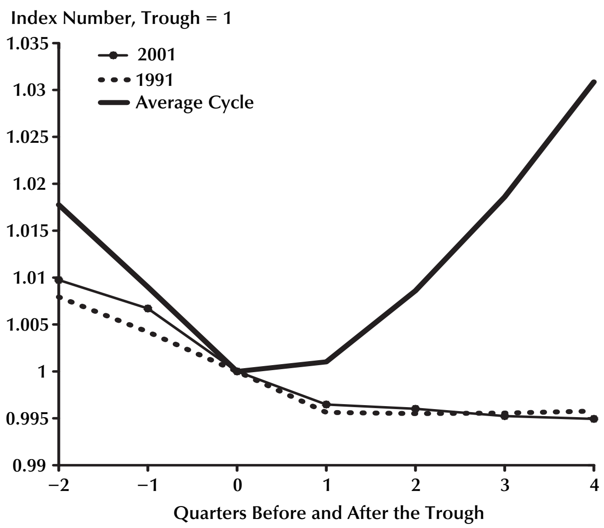
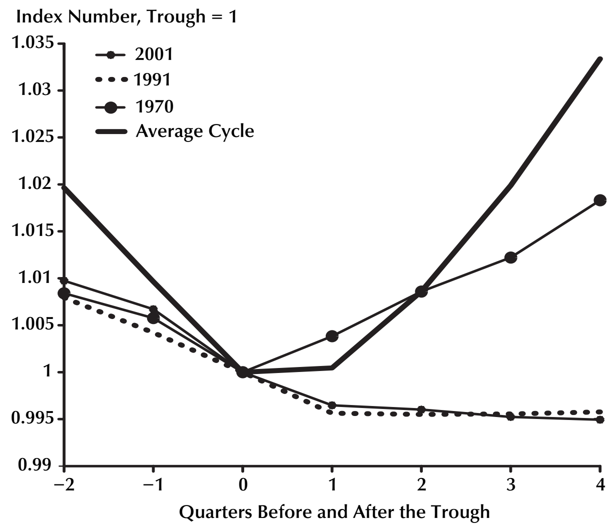


Figure 4

The 1970 Recovery (Schreft-Singh Method)



in their paper. Whereas the typical recovery shows steadily rising employment, with an increase of more than 3 percent in the first year of the recovery, the two most recent recessions show employment decreasing in each of the four quarters following the turning point. While this picture certainly suggests that the two recent recoveries are different from the average of the preceding ones, it obviously does not tell us whether there are previous episodes that also resemble the two recent ones. For our purposes we are particularly interested in whether the recovery that began in 1970 also displays this pattern. Figure 4 repeats the analysis of Figure 3 except we now consider three recoveries individually and compare them with the average of the remaining five recoveries. (Throughout this analysis we ignore the recovery in the early 1980s because it was so short-lived.)

This figure suggests that the recovery that began in 1970 is much more similar to the average recovery than it is to the recoveries following the two most recent recessions. One can repeat this analysis for the other recoveries as well, and one obtains a similar pattern in each case. Based on this analysis, one would be led to conclude that it is only the two most recent recoveries that have had particularly distinctive employment dynamics.

However, there are several issues we want to raise regarding the Schreft-Singh method of summarizing the data. The first issue is that the Schreft-Singh method is not consistent with modern views of the business cycle. Following Lucas (1977), modern business cycle analysis views the business cycle as deviations from a slowly changing trend. Properties of business cycles should be properties of the component of the time series that corresponds to these deviations from trend. The Schreft-Singh method neglects this consideration in two important regards. First, some recessions are more severe than others. To the extent that recessions are temporary departures from trend, a deeper recession would naturally be expected to be followed by higher subsequent growth in employment. The Schreft-Singh method does not incorporate this feature. Second, their method does not distinguish between movements in the trend and deviations from the trend. If the (raw) level of employment following the trough of a recession starts to increase, how are we to know to what extent we are moving closer to trend? If the trend always increased at the same rate,

Table 1

Employment Growth by Decade (percent)

1950-60	1960-70	1970-80	1980-90	1990-2000
20.89	26.81	23.91	17.85	17.45

this issue would be irrelevant because it would affect all recoveries in the same fashion. However, a key feature of the postwar labor market in the United States is that trend employment growth has fluctuated substantially over time, due both to the entry of the baby boom into the labor market and the increased participation of women. Table 1 illustrates this point by showing the decadal growth rates in employment for the U.S. economy for the five postwar decades.

Table 1 shows that the differences are large: The decadal growth rate in employment during the 1960s is more than one and a half times as large as the decadal growth rates in the two most recent decades. It follows that sorting out relative movements in trend and deviation from trend may be an important consideration in documenting the differential pace of employment growth during recoveries.

Third, the Schreft-Singh method compares the dynamics of recoveries by examining the behavior going forward from the turning point. It is not clear that the turning point is the appropriate comparison point across cycles. In particular, if the downturns preceding the recoveries have been different, it is not clear that behavior should be the same from the turning point forward. In fact, we will argue later that our model suggests that the turning point should not be used as a common reference point.

Having raised some issues about the statistics that Schreft and Singh report, we now describe the method that we use.

An Alternative Look at the Data

Our method is straightforward and is consistent with current practice in business cycle analysis in terms of documenting properties of cyclical fluctuations. In particular, let X_t be a quarterly series that is seasonally adjusted for which we have observations going from period 0 to period N. Define x_t to be the log of the series X_t. We define the trend component of x_t, denoted by x_t^T, by using the Hodrick-Prescott filter. In particular, x_t^T is the solution to the following optimization problem:

$$\min_{\{x_t^T\}} \sum_{t=1}^{N-1} \left\{ \left[x_t - x_t^T \right]^2 + \lambda \left[\left(x_{t+1}^T - x_t^T \right) - \left(x_t^T - x_{t-1}^T \right) \right]^2 \right\}.$$

Following the literature, for quarterly data we use a value of $\lambda = 1{,}600$.

The cyclical component of x_t, denoted by x_t^C, is simply the deviation of x_t from its trend value: $x_t^C = x_t - x_t^T$. Because the series are measured in logs, the cyclical component reflects the percent deviation of the variable from its trend.

Figure 5 repeats the exercise of Schreft and Singh but uses the cyclical component defined previously as opposed to the raw data. In particular, this graph shows the percent change in the cyclical component of employment in each of the four quarters following the NBER turning points.[11]

This figure tells a similar story to the one told by Figure 3, though we note that some details are different. In particular, whereas Figure 3 indicated that in a typical recovery employment begins to grow as soon as the turning point is reached, Figure 5 displays the well-known feature that employment lags gross domestic product (GDP). In particular, this figure shows that in a typical recovery, employment begins to increase one quarter after GDP begins to increase.[12] (We note that the cyclical component of GDP defined here has the property that the turning point for GDP follow-

[11] It is important to note that, whenever one detrends the data, the behavior of the cyclical component at the very beginning and end of the sample is somewhat sensitive to the initial and terminal data points. This implies that the properties of the 2001 recovery may look somewhat different as more data become available.

[12] See, for example, the cyclical properties as reported in Cooley and Prescott (1991).

Figure 5

The 1991 and 2001 Recoveries (Detrended)

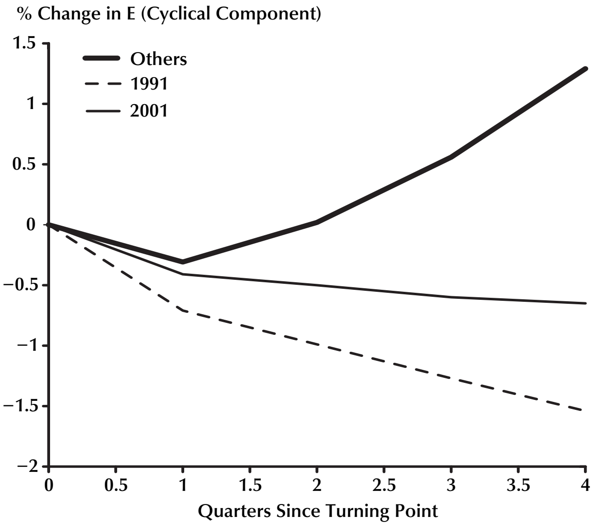
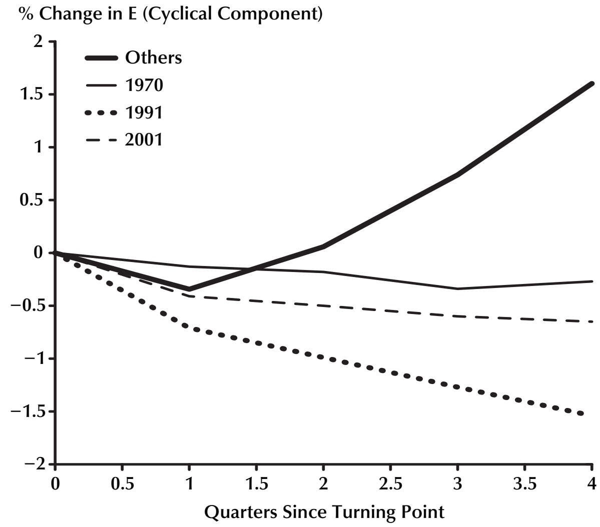


Figure 6

Recoveries of 1970, 1991, and 2001 (Detrended)



ing each recession coincides with the NBER turning point dates.)

Next, we again ask if the recovery that began in 1970 is similar to the two most recent recoveries. Figure 6 shows the results of carrying out the exercise analogous to that which generated Figure 4—that is, we now consider three individual recoveries and compare them with the average of the other five.

We see that Figure 6 offers a very different conclusion than does Figure 4. Based on analysis of the cyclical component of the employment series, the 1970 recovery also shows that, even one year after the turning point, employment remains below its level at the turning point. While this picture still indicates quantitative differences across the three recoveries, the qualitative behavior is in fact similar.

These figures still suffer from the first problem we mentioned earlier: They do not take into account that the recessions vary quite substantially in their severity, and hence we do not know how much of the variation in growth simply reflects differences in distances from trend. There are many ways that we might normalize business cycles to account for the differing magnitudes. We employ a simple procedure here, which is to normalize by the magnitude of the recession as measured by the maximum percent deviation of output below trend. We then scale all of the employment deviations by dividing through by the absolute value of this number. This calculation is shown in Figure 7.

Qualitatively this figure presents the same conclusions as the earlier figure, but we note that it does affect the quantitative differences across episodes.

One way to summarize the properties of the figures is in terms of the extent to which the turning point for employment lags the turning point for GDP. Table 2 shows the values for each of the post-1950 recessions.

This statistic confirms a property that we saw in the figures presented earlier: Whereas a lag of 0 or 1 quarter is typical for the postwar period, there are three recoveries in which the lag is longer, and all three of these are recoveries from recessions that follow long expansions. This evidence supports our earlier claim about characterizing postwar recoveries. At the same time, Table 2 suggests that there is quite a significant difference between the 1970 recovery and the two most

Figure 7

Cyclical Employment, Normalized by Magnitude of Trough

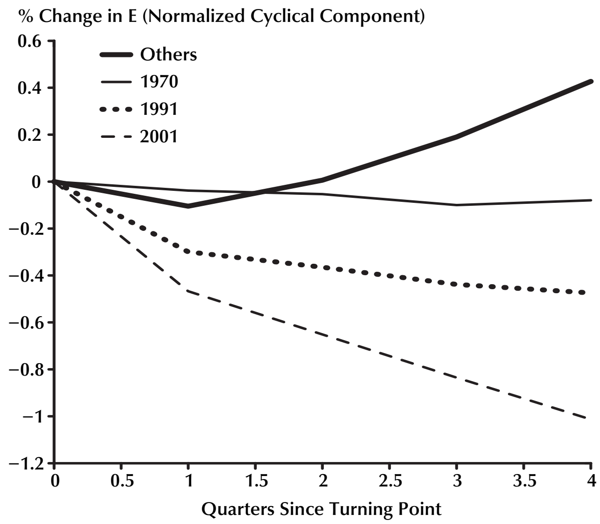


Figure 8

Comparing the 1970, 1991, and 2001 Recoveries

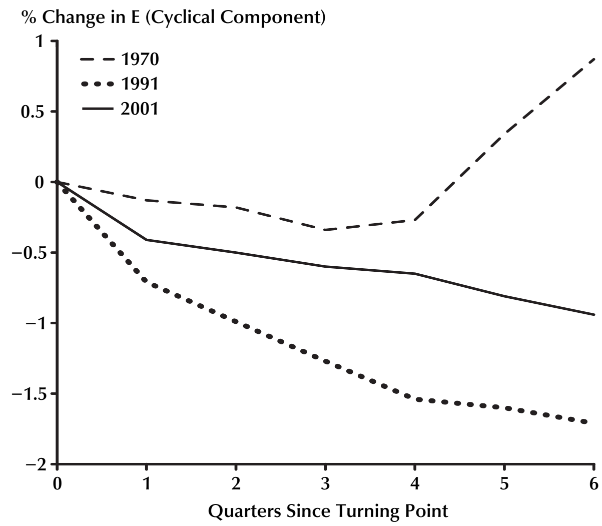


Table 2

Lag in Employment Turning Point Relative to GDP (quarters)

1954	1958	1961	1970	1975	1982	1991	2001
1	0	1	3	1	1	6	7

recent recoveries. The lags in the two most recent recoveries are in fact much longer than was experienced in the 1970 recovery. In particular, if we extend the analysis from the four quarters following the turning point to seven quarters, then we obtain Figure 8.

Figure 8 would seem to suggest that the 1970 episode is not so similar to the two most recent ones. What we argue next, however, is that this difference is illusory. In particular, we argue that it is driven by the choice of initial point and that, when this initial point is chosen in a way that we believe is more consistent with the theory laid out earlier in the paper, the differences disappear.

To understand the issue, consider the discussion from the previous section. The key point there was that the time before employment begins to increase is tied to the time required for reorganization to diminish sufficiently. The key point to take away from this analysis is that the turning point is not the point at which reorganization begins. Presumably reorganization starts to take place at some point during the recession preceding the recovery. This possibility is significant because the duration of the recession preceding the turning point differs significantly across recessions. In particular, this period is either average or below average for the two most recent recessions and is much above average for the 1969-70 recession.

To implement this element, we need to have a method for picking out at what point reorganization begins. One possibility would be that it begins when the downturn first begins. However,

Table 3

Quarters Before Turning Point with GDP Below Trend

1954	1958	1961	1970	1975	1982	1991	2001
2	2	2	4	2	4	1	2

Figure 9

Recoveries with Alternative Initial Point

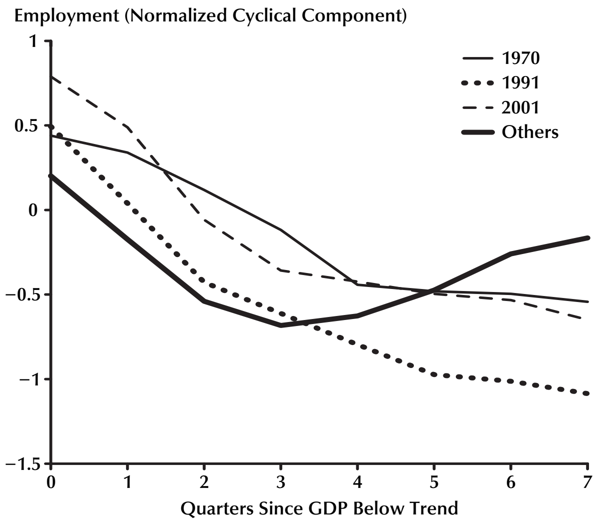
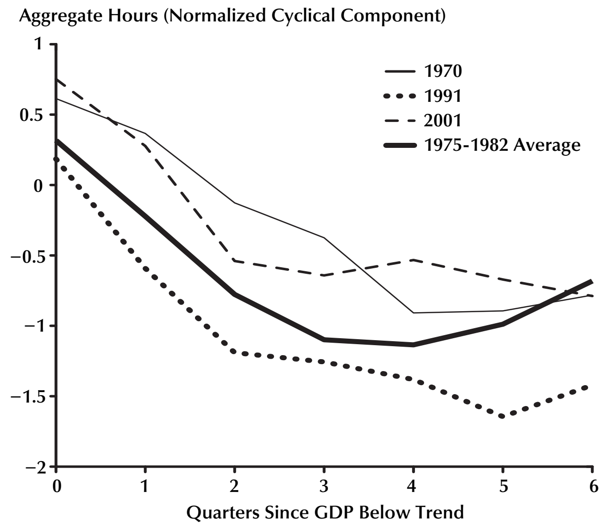


Figure 10

Behavior of Aggregate Hours



when a recession first begins, output is still quite high above trend and the analytic result that we proved earlier in the paper suggested that it is the level of output that is particularly important. With this in mind, we identify the point at which reorganization begins to be the first quarter in which the cyclical component of output lies below trend. This choice is obviously ad hoc, so the analysis that follows should really be interpreted as illustrating the potential importance of this type of correction. Table 3 shows, with the data, how many quarters this occurs prior to the NBER turning point for each of the eight recessions.

Table 3 indicates that the typical number of quarters that GDP is below trend prior to reaching its turning point is two, but that in 1970 this value was four, and in 1991 it was only one.

What we do next is to repeat our earlier analysis, but instead of using the turning point as the initial condition for each recession, we use the period as indicated in Table 3. Once again we continue to normalize using the magnitude of the drop in GDP as our scale factor. Figure 9 shows the average of the five "typical" recessions versus each of the three more prolonged recoveries.

It shows that the three recessions all stand out as different from the average of the others.

To this point we have focused on employment dynamics. Employment is simply one dimension along which labor input can vary. An important issue is the extent to which different behavior of employment across cyclical episodes also represents differences in labor input. Alternatively, it could be that the dominant difference across episodes is the compositional changes in labor input. More generally, one should also be con-

Figure 11

Productivity per Employee in the 1970, 1991, and 2001 Recoveries

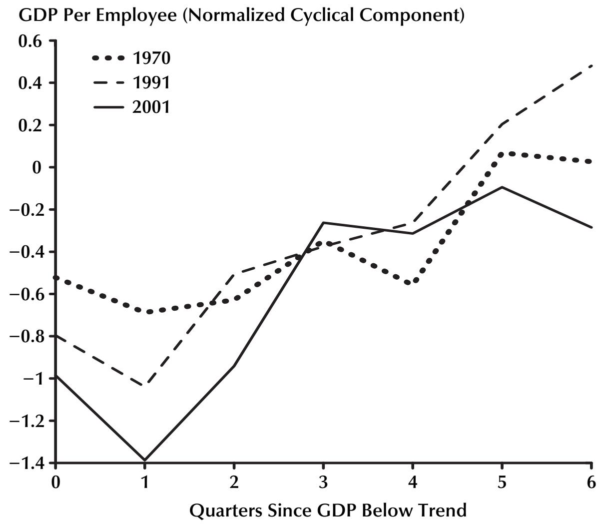
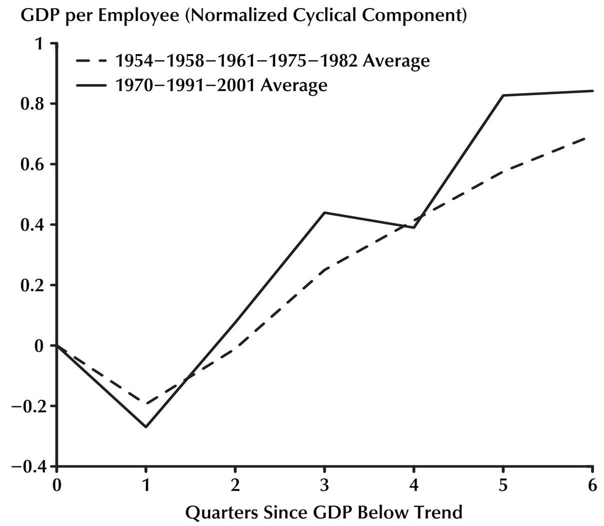


Figure 12

Comparison of Productivity per Employee Across Recoveries



cerned about measuring effective labor input rather than simply bodies or hours. We leave a more careful analysis of these issues for future work, but do want to present at least one piece of evidence to suggest that the differences in the behavior of employment that we have noted also extend to the behavior of aggregate hours. We have hours data only from 1964 on, so this analysis has only five cyclical episodes to compare. Figure 10 compares the three recoveries that we have focused on with the average of the other two recoveries. Period 0 in this figure refers to the first period in which GDP drops below trend prior to the associated recovery.

While the basic pattern in this figure is qualitatively similar to the one found in our graphs for employment, it is also true that the differences are much smaller in this graph than in the employment graphs. We infer from this that a more careful study of the behavior of labor input along the intensive and extensive margins is warranted.

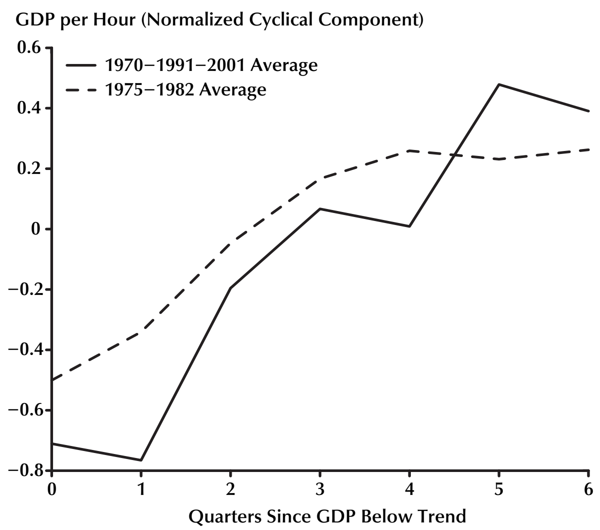
Last, it is of interest to examine the behavior of productivity across the cyclical episodes. Our model predicts that when organizations switch from producing to reorganizing they experience a decrease in productivity. However, as time passes and successful reorganization occurs, we should see increases in productivity. Having said this, we also feel that a large degree of caution need be taken with respect to assessing the implications for productivity. We have implicitly assumed that aggregate fluctuations in our model are driven by shocks to the demand for the output of each organization. While this is a convenient formulation for our analysis, it could be that the increase in demand is driven by improvements in product quality, which in a more complete model would also show up as productivity changes.

Having offered this caveat, we now turn to analyze the dynamics of productivity. Because hours data are available only since 1964, we use two different measures of productivity. First we compare output per worker to be able to use all eight recessions and then use output per hour to compare the five most recent recessions. Figure 11 shows how productivity per worker evolves in the three recoveries of interest.

Note that in all cases productivity drops in the initial period but then increases thereafter.

Figure 13

Comparison of Productivity per Hour Across Recoveries



While the magnitudes are somewhat different across the three episodes, the pattern is quite similar. At a very qualitative level this pattern seems to accord well with the implications of our model described previously. Next we consider how productivity changes vary across the different types of recoveries. Figure 12 compares the average of these three recoveries with the average across the other five recoveries.

Both curves show the same qualitative behavior. The average of the three recoveries following long expansions does indeed have a slightly larger drop in the initial period and does show somewhat higher subsequent growth. Qualitatively, these patterns are consistent with what one would expect from our model, though the quantitative differences do not seem that large.

Next we compare the behavior of productivity per hour for the post-1964 recoveries.

Figure 13 shows the same two features: The three recoveries associated with recession following long expansions have a somewhat larger initial drop in productivity and subsequently experience somewhat higher growth.

CONCLUSION

We have highlighted a simple economic mechanism that we argue may be relevant for understanding the different behavior of labor market aggregates across business cycles. The model stresses two key effects. First, it argues that internal organizational dynamics are affected by aggregate shocks. Second, it stresses that the situations of organizations affect the manner in which the economy responds to aggregate shocks. In periods of high economic activity, organizations postpone structural changes to take advantage of current opportunities. But once an organization begins the process of restructuring, it is less likely to hire workers and more likely to release workers. These effects suggest the possibility that long expansions will be followed by recoveries in which employment starts to increase much later than output.

We then assess this link by studying eight U.S. recessions in the post-1950 period. We argue that all three recoveries from recessions that followed long expansions exhibit the pattern of a long delay in the turning point for aggregate employment. This finding contrasts sharply with the characterization that it is the two most recent recoveries that are distinct.

While we think this work is suggestive, we must also emphasize that it is indeed only suggestive. A more rigorous quantitative assessment of the economic mechanism is called for, as is a more thorough analysis of the data.

REFERENCES

Aaronson, Daniel; Rissman, Ellen R. and Sullivan, Daniel G. "Further Evidence on Sectoral Reallocation and the Jobless Recovery. " Unpublished manuscript, Federal Reserve Bank of Chicago, 2004.

Andolfatto, David and MacDonald, Glenn M. "Jobless Recoveries." Working paper, 2004.

Bertschek, Irene and Kaiser, Ulich. "Productivity Effects of Organizational Change: Microeconometric Evidence." Unpublished manuscript, 2001.

Burns, Arthur F. and Mitchell, Wesley C. *Measuring*

Business Cycles. New York: National Bureau of Economic Research, 1946.

Caballero, Richard and Engel, Eduardo. "Explaining Investment Dynamics in U.S. Manufacturing: A Generalized (S,s) Approach." *Econometrica*, July 1999, *67*(4), pp. 783-826.

Cole, Harold L. and Ohanian, Lee E. "New Deal Policies and the Persistence of the Great Depression: A General Equilibrium Analysis." *Journal of Political Economy*, August 2004, *112*(4), pp. 779-816.

Cooley, Thomas F. and Prescott, Edward C. "Economic Growth and Business Cycles," in Thomas F. Cooley, ed., *Frontiers of Business Cycle Research*. Princeton, NJ: Princeton University Press, 1991, pp. 1-38.

Groshen, Erica L. and Potter, Simon M. "Has Structural Change Contributed to a Jobless Recovery?" Federal Reserve Bank of New York *Current Issues in Economics and Finance*, August 2003, *9*(8).

Hall, Robert E. "Labor Demand, Labor Supply and Employment Volatility," in Olivier Blanchard and Stanley Fischer, eds., *NBER Macroeconomics Annual 1991*. Cambridge, MA: MIT Press, 1991, pp. 17-61.

Hopenhayn, Hugo A. "Entry, Exit and Firm Dynamics in the Long Run." *Econometrica*, September 1992, *60*(5), pp. 1127-50.

Hopenhayn, Hugo A. and Rogerson, Richard. "Job Turnover and Policy Evaluation: A General Equilibrium Analysis." *Journal of Political Economy*, October 1993, *101*(5), pp. 915-38.

Kliesen, Kevin L. "The 2001 Recession: How Was It Different and What Development May Have Caused It?" Federal Reserve Bank of St. Louis *Review*, September/October 2003, *85*(5), pp. 23-38.

Koenders, Kathryn. "Long Expansions and Slow Recoveries: A Closer Look at Employment Fluctuations." Unpublished manuscript, Arizona State University, 2005.

Krusell, Per and Smith, Anthony A. "Income and Wealth Heterogeneity in the Macroeconomy." *Journal of Political Economy*, October 1998, *106*(5), pp. 867-96.

Lilien, David M. "Sectoral Shifts and Cyclical Unemployment." *Journal of Political Economy*, August 1982, *90*(4), pp. 777-93.

Lucas, Robert E. Jr. "Understanding Business Cycles." *Carnegie-Rochester Series on Public Policy*, 1977, *5*, pp. 7-29.

Lucas, Robert E. Jr. "On the Size Distribution of Business Firm." *Bell Journal of Economics*, Autumn 1978, *9*(2), pp. 508-23.

Schreft, Stacey L. and Singh, Aarti. "A Closer Look at Jobless Recoveries." Federal Reserve Bank of Kansas City *Economic Review*, Second Quarter 2003, *88*(2), pp. 45-73.

Shimer, Robert. "The Cyclicality of Hires, Separations, and Job-to-Job Transitions." Federal Reserve Bank of St. Louis *Review*, July/August 2005a, *87*(4), pp. 493-507.

Shimer, Robert. "Reassessing the Ins and Outs of Unemployment." Working paper, 2005b.

Thomas, Julia J. "Is Lumpy Investment Relevant for the Business Cycle?" *Journal of Political Economy*, June 2002, *110*(3), pp. 508-34.

van Rens, Thijs. "Organizational Capital and Employment Fluctuations." Job Market Paper, Princeton University, 2004.

Veracierto, Marcelo L. "Plant Level Irreversible Investment and the Business Cycle." *American Economic Review*, March 2002, *92*(1), pp. 181-97.

Commentary

Fernando Alvarez

I would like to start my discussion with some quotes from Ben Bernanke's (2003) speech about the behavior of employment during the past recovery. In this speech Bernanke thoroughly reviews the main features of the data and several of the hypotheses advanced to understand the sluggishness of employment during the recovery.

> I suspect that some of the recent expansion in productivity is instead the delayed result of firms' heavy investment in high technology equipment in the later part of the 1990s. Only over time have managers learned how to organize their production and distribution so as to take full advantage of these technologies... [Productivity growth] has also enabled firms to meet the demand for their output without hiring workers.
> ...Although other explanations for the jobless recovery...have played a role, in my view the productivity explanation is, quantitatively, probably the most important.

Bernanke's interpretation, while not identical to the one in the paper by Koenders and Rogerson (2005), is very close: It emphasizes technology as the driving force, as opposed to aggregate demand fluctuations, and locates the seed of the sluggishness of the recovery in the strength of the previous expansion.

My discussion centers on a presentation of a general equilibrium version of the model by Koenders and Rogerson. Nevertheless, before offering it, I would like to briefly mention three other points.

First, as it is well known, the magnitude of the decline in employment during and after the past recession differs according to the data sets used to measure it. The establishment survey, which is the one used in the paper, shows a much larger decline than the household survey does. There are several differences in the design of the two surveys, which can, in principle, account for the difference in employment. The establishment survey measures employment from payroll information from a large number of firms, so it gives an estimate of the total employment for a large sector of the U.S. economy. The household survey measures the employment status from a given set of households, so it gives an estimate of the employment/population ratio. One explanation that has been advanced for the difference between the two estimates is that, if population decreases, it is possible that employment/population stays constant and that total employment decreases. Because population is measured accurately only every decade, a decline in population may explain part of the difference between the two estimates. Indeed, it has been proposed that immigration may have sharply declined during the recession and the beginning of the recovery, as a consequence of policies enacted after September 11, 2001. The paper by Koenders and Rogerson uses Hodrick-Prescott (HP) filtered data for employment, which, in principle, should remove slow-moving variations (such as the one in population) from employment. Nevertheless, it is well known that the HP filter is much less accurate at the end points, so one may be suspicious of its ability to remove a

Fernando Alvarez is a professor of economics at the University of Chicago.

Federal Reserve Bank of St. Louis *Review*, July/August 2005, *87*(4), pp. 581-87.

change in the trend that happened in the past three years of the sample. Additionally, the models we typically write are in terms of employment/population ratios, so it seems that, even though the household survey may be noisier, its results should be used or at least compared with the ones for the establishment survey used in the paper.

My second comment is that the model focuses on the (time-varying) cyclicality of firing. Nevertheless, recent studies by R. Shimer and R. Hall document that firing is less cyclical than hiring. While the paper recognizes this possibility, the model is still one in which the variations are mostly in firing.

My third point is related to the measurement of the hypothesis of the paper: The strength of the recovery after a recession depends on the length of the expansion that precedes it. There are only so many recessions in the United States after World War II. Thus, I think it may be useful to repeat the empirical exercise for the Organisation for Economic Co-operation and Development (OECD) countries. Of course, business cycles across these countries are correlated, but not perfectly so; thus, I think the international comparison will add valuable information to the hypothesis.

A GENERAL EQUILIBRIUM VERSION OF THE KOENDERS AND ROGERSON MODEL

Here I describe a general equilibrium (GE) version of the Koenders and Rogerson model (KR hereafter). This GE version endogenizes wages and interest rates. I first describe a "reduced form" model, which I hope makes the main ingredients of the theory easy to see. Then I show that the reduced form is the outcome of a GE version of the KR model.

Although the preferences and technology are similar, there are two main differences between the KR model and the model here. One difference is that I consider a planner's problem instead of a market economy. The second difference is that I allow heterogeneity in the cost of reorganization, as explained in the next section. In the following description, I keep their notation as much as possible.

The way I converted their model into a GE problem is by postulating preferences of the form

$$\sum_{t=0}^{\infty} \beta^t E_0 \left(\frac{\left(C_t\right)^{1-\gamma}}{1-\gamma} - \left(\frac{A}{\varepsilon_{2t}}\right) n_t \right),$$

where C_t is aggregate consumption and n_t is employment. An increase in ε_2 makes the household less willing to work, and hence they will, everything else equal, end up producing less and consuming less. In this sense, ε_2 is a "demand shock." As in KR, there is a constant number of firms in the economy. Let μ_{2t} denote the beginning period fraction of non-reorganized firms and μ_2' the next period value of this variable. The more non-reorganized firms are in the economy, the fewer reorganized firms are. A non-reorganized firm incurs an extra fixed cost of production relative to reorganized firms. A non-reorganized firm can become reorganized by incurring a costly action that involves producing with a higher marginal cost. The technological possibilities for this economy are described by

$$n_t = a\left(\mu_{2t}, \mu_{2t+1}\right) C_t + \mu_{2t}\phi,$$

where n_t is the labor units required to produce C_t units of the aggregate consumption if the number of non-reorganized firms today is μ_{2t} and its number next period is μ_{2t+1}. The parameter $\phi > 0$ measures the fixed cost, in terms of labor, that each non-reorganized firm incurs. The function a gives the marginal cost, in terms of labor, of producing each unit of C_t. It is assumed that $a(\mu_{2t}, \mu_{2t+1})$ is decreasing in μ_{2t+1} because reorganization is costly; lowering μ_{2t+1} saves on the current costly action of reorganization. It is assumed that $a(\mu_{2t}, \mu_{2t+1})$ is increasing in μ_{2t}, because non-reorganized firms are less productive. The next section derives these assumptions from a GE model that uses the same technology as the partial equilibrium KR model. Having described the preferences and technology I consider the following *reduced form planning problem*:

$$V\left(\mu_{2t},\varepsilon_2\right)$$
$$=\max_{C,n,\mu_2'}\left\{\frac{C^{1-\gamma}-1}{1-\gamma}-\left(A/\varepsilon_2\right)n+\beta E\left[V\left(\mu_2',\varepsilon_2'\right)\middle|\varepsilon_2\right]\right\}$$
$$n=a\left(\mu_2,\mu_2'\right)C+\phi\mu_2.$$

The optimal decision rules are given by functions g and n:

$$\mu_{2t+1}=g\left(\mu_{2t},\varepsilon_{2t}\right)$$
$$n_t=n\left(\mu_{2t},\varepsilon_{2t}\right).$$

For the case of log utility ($\gamma=1$) I obtain the following: g is increasing in μ_2 with a slope less than 1, so that the dynamics are stable; it is increasing in ε_2, so at times of "high demand" there is no reorganization and thus μ_2 increases more. n is given by

$$n_t=n\left(\varepsilon_{2t},\mu_{2t}\right)=\frac{\varepsilon_{2t}}{A}+\phi\mu_{2t}.$$

This decision rule balances two effects: Employment increases with the "demand" shock, ε_{2t} (the household is more willing to work and consume), and also increases with μ_{2t} because the reorganized firms have larger fixed costs.

I can now reproduce the main result in their paper, explained in section 4.1. Consider an initial value of μ_{20} that is low, reflecting that a long expansion has occurred up to now. Consider an initial condition for ε_{20} low enough, so there will be reorganization and a subsequent increasing sequence of ε_{2t}, which is meant to imitate the impulse response of a persistent shock. Then, μ_{2t} will be increasing and hence the path of employment given by the optimal decision rule just described will be u-shaped, with its lower point after the lower point of ε_{2t}—hence, the sluggish recovery after a long expansion.

One point that differs between the GE and the partial equilibrium versions is the characterization of the conditions under which reorganization does occurs in relation to the persistence of the demand shock ε_2. In the GE version of the model with $\gamma=1$, if ε_2 is i.i.d., then g is independent of ε_2. This differs from the partial equilibrium version. But for the case that is arguably more interesting empirically, they are similar. In the GE version with $\gamma=1$, the decision rule g is increasing in ε_2 if the distribution of ε_{2t+1} is stochastically higher when ε_{2t} is larger. In other words, in the case of persistent "demand shocks," reorganization is countercyclical. Hence, the GE version of the model reproduces the key results sought by KR.

DETAILED DESCRIPTION OF THE GE MODEL

I first describe the technology and preferences of the GE model and then show that they imply the reduced form introduced here—the a' function—and that the optimal decision rules have the properties previously described.

I introduce heterogeneity in the reorganization cost, $\bar{\eta}$, as a way to smooth out the problem. In each period there is a continuum of goods indexed by $i\in[0,\bar{\mu}]$. These goods correspond to three types of firms. There are μ_i firms of each type for $i=1,2,3$. Type 1 goods are those of firms with low idiosyncratic demand shock ε^s and labor requirement a^s. Type 3 goods corresponds to reorganized large firms with high idiosyncratic demand shock ε^l and labor requirement a^l. Type 2 goods correspond to firms with high idiosyncratic demand shock ε^l and labor requirements that depend on whether they are reorganizing or not and on the idiosyncratic reorganization cost, $\bar{\eta}\in R_+$, assumed to be drawn i.i.d. each period from a distribution with cdf G. I anticipate that the form for optimal policy is that those firms with $\bar{\eta}<\bar{\eta}^*$ will engage in "organization restructuring" and those with cost $\bar{\eta}>\bar{\eta}^*$ will not. I denote the output of those that do reorganize as $y_2(\bar{\eta})$ and the common level of output of those that do not reorganize as $\bar{y}_2$. The labor requirements of each type of firm are

$$n_1=a^s y_1,$$
$$n_2\left(\bar{\eta}\right)=a^l\left(1+\eta\right)y_2\left(\bar{\eta}\right)+\phi \text{ for } \bar{\eta}\le\bar{\eta}^*,$$
$$n_2\left(\bar{\eta}\right)=a^l\bar{y}_2+\phi \text{ for } \bar{\eta}>\bar{\eta}^*,$$
$$n_3=a^l y_3.$$

This gives a total labor requirement equal to

(1)

$$n = \mu_1\left(a^s y_1\right) + \mu_2\left[a^l \int_0^{\bar{\eta}}\left[\left(1+\bar{\eta}\right) y_2\left(\bar{\eta}\right)\right] dG\left(\bar{\eta}\right) + a^l\left(1-G\left(\bar{\eta}^*\right)\right)\bar{y}_2\right] + \mu_2\phi + \mu_3\left(a^l y_3\right).$$

The law of motion of the number of firms of each type is as follows. Each period, measure 1 of firms of type 1 enter, fraction λ^l of the existing firms die, and fraction π^e become type 2 firms, so

$$\mu_2' = \mu_1\left(1-\lambda^s-\pi^l\right)+1.$$

Fraction λ^l of type 2 firms die, and fraction $G(\bar{\eta}^*)$ engage in restructuring, with π^e of those restructuring doing it successfully, so

$$\mu_2' = \left(1-\lambda^l-G\left(\bar{\eta}^*\right)\pi^e\right)\mu_2 + \pi^l\mu_1.$$

Fraction λ^l of type 3 firms die, so

$$\mu_3' = \left(1-\lambda^l\right)\mu_3 + G\left(\bar{\eta}^*\right)\pi^e\mu_2 .$$

Letting

$$\bar{\mu} \equiv \frac{1+\pi^l/\lambda^l}{\lambda^s+\pi^l},$$

one can verify that if

$$\mu_{10} = \frac{1}{\lambda^s+\pi^l}\left(\equiv\bar{\mu}_1\right)$$

and

$$\mu_{10}+\mu_{20}+\mu_{30} = \bar{\mu},$$

then

$$\begin{aligned}\mu_{1t} &= \bar{\mu}_1\\ \mu_{1t}+\mu_{2t}+\mu_{3t} &= \bar{\mu}\end{aligned}$$

for all $t \geq 0$. Thus, without loss of generality, the number of type 1 firms is constant ($= \bar{\mu}_1$), as well as the total number of firms ($\bar{\mu}$), so that

(2)
$$\mu_3 = \bar{\mu} - \bar{\mu}_1 - \mu_2$$

and the law of motion of firms of type 2 is

(3)
$$\mu_2' = \left(1-\lambda^l-G\left(\bar{\eta}^*\right)\pi^e\right)\mu_2 + \pi^l\bar{\mu}.$$

Consumption, C_t, is a Dixit-Stiglitz aggregate of a continuum of goods:

$$C_t = \left[\int_0^{\bar{\mu}}\alpha_i\left[c_{it}\right]^{\frac{\varphi-1}{\varphi}}di\right]^{\frac{\varphi}{\varphi-1}},$$

where $\alpha_i > 0$ are the weights of the different goods. The weight of each type is as follows: Type 1 has weight $\alpha_i = \varepsilon_1$, and type 2 and 3 goods have weights $\alpha_i - \varepsilon^l$. This captures the higher demand for type 1 and 2 firms in the partial equilibrium model. Hence,

(4)

$$C_t = \left[\int_0^{\bar{\mu}}\alpha_i\left[c_{it}\right]^{\frac{\varphi-1}{\varphi}}di\right]^{\frac{\varphi}{\varphi-1}} = \left[\begin{array}{l}\left(\mu_1\varepsilon^s\right)\left(y_{1t}\right)^{\frac{\varphi-1}{\varphi}}\\ +\left(\mu_2\varepsilon^l\right)\left(\begin{array}{l}\int_0^{\bar{\eta}^*}\left[y_{1t}\left(\bar{\eta}\right)\right]^{\frac{\varphi-1}{\varphi}}dG\left(\bar{\eta}\right)\\ \quad+\left[\bar{y}_{2t}\right]^{\frac{\varphi-1}{\varphi}}\left(1-G\left(\bar{\eta}^*\right)\right)\end{array}\right)\\ +\left(\mu_3\varepsilon^l\right)\left(y_{3t}\right)^{\frac{\varphi-1}{\varphi}}\end{array}\right]^{\frac{\varphi}{\varphi-1}}.$$

Now I can write the *planning problem* as

$$V\left(\mu_2,\varepsilon_2\right) = \max_{C,y,\eta^*}\left\{\frac{\left[C\right]^{1-\gamma}}{1-\gamma} - \left(A/\varepsilon_2\right)n + \beta E\left[V\left(\mu_2',\varepsilon_2'\right)\middle|\varepsilon_2\right]\right\},$$

where y is the vector of output across firms,

$$y = \left(y_1, y_2\left(\bar{\eta}\right)_{\bar{\eta}=0}^{\bar{\eta}=\bar{\eta}^*}, \bar{y}_2, y_3\right),$$

and where the maximization is subject to the definition of the aggregate consumption C (4), the total labor requirement n (1), the implied value for μ_3 (2), and the law of motion for μ_2 (3).

ANALYSIS OF THE GE MODEL

First I turn to the derivation of the reduced form planning problem, and then I analyze the optimal decision rules of the reduced form planning problem.

Let $e(\mu_2,\mu_2')$ be the value of cutoff point $\bar{\eta}^*$ so that the law of motion implies μ_2' given the current μ_2. The function e solves

$$\mu_2' = \left(1 - \lambda^l - G\left(e\left(\mu_2', \mu_2\right)\right)\pi^e\right)\mu_2 + \pi^l \bar{\mu}.$$

For future reference, notice that

$$\frac{\partial e}{\partial \mu_2} = \frac{\left(1 - \lambda^1 - G\pi^e\right)}{G'\pi^e \mu_2} > 0,$$
$$\frac{\partial e}{\partial \mu_2'} = -\frac{1}{\pi^e G' \mu_2} < 0.$$

Define $\tilde{a}(\mu_2, \bar{\eta}^*)$ as the minimum employment n needed to produce one unit of Dixit-Stiglitz C, formally:

$$\tilde{a}\left(\mu_2, \bar{\eta}^*\right)C \equiv \min_{y_1, y_2, y_3} \left\{ \begin{array}{l} \bar{\mu}_1 a^s y_1 + \mu_2 a^l \left(\begin{array}{l} \int_0^{\bar{\eta}^*} \left(1+\bar{\eta}\right) y_2\left(\bar{\eta}\right) dG \\ + \bar{y}_2\left(1 - G\left(\bar{\eta}^*\right)\right) \end{array} \right) \\ + \mu_2 \phi + \mu_3 a^l y_3 \end{array} \right\}$$

subject to

$$C = \left[\begin{array}{l} \bar{\mu}_1 \varepsilon^s y_1^{\frac{\varphi-1}{\varphi}} \\ + \mu_2 \varepsilon^l \left(\int_0^{\bar{\eta}^*} y_2\left(\bar{\eta}\right)^{\frac{\varphi-1}{\varphi}} dG + \bar{y}_2^{\frac{\varphi-1}{\varphi}} \left(1 - G\left(\bar{\eta}^*\right)\right) \right) \\ + \mu_3 \varepsilon^l y_3^{\frac{\varphi-1}{\varphi}} \end{array} \right]^{\frac{\varphi}{\varphi-1}}.$$

Using the first-order conditions of this minimization problem into the objective function, it can be shown that

$$\tilde{a}\left(\mu_2, \bar{\eta}^*\right) = \left[\begin{array}{l} \bar{\mu}_1 \varepsilon^s \left(\frac{a^s}{\varepsilon^s}\right)^{1-\varphi} \\ + \mu_2 \varepsilon^l \left(\frac{a^l}{\varepsilon^l}\right)^{1-\varphi} \left(\begin{array}{l} \int_0^{\bar{\eta}^*} \left(1+\bar{\eta}\right)^{1-\varphi} dG\left(\bar{\eta}\right) \\ + \left(1 - G\left(\bar{\eta}^*\right)\right) \end{array} \right) \\ + \left[\bar{\mu} - \bar{\mu}_1 - \mu_2\right] \varepsilon^L \left(\frac{a^l}{\varepsilon^l}\right)^{1-\varphi} \end{array} \right]^{\frac{1}{1-\varphi}}.$$

For future reference notice that

$$\frac{\partial \tilde{a}\left(\mu_2, \bar{\eta}^*\right)}{\partial \mu_2} = \frac{\tilde{a}^{-\varphi}}{1-\varphi} \left(\frac{a^l}{\varepsilon^l}\right)^{1-\varphi} \varepsilon^l \left\{ \left[\begin{array}{l} \int_0^{\bar{\eta}^*} \left(1+\bar{\eta}\right)^{1-\varphi} dG\left(\bar{\eta}\right) \\ + \left(1 - G\left(\bar{\eta}^*\right)\right) \end{array} \right] - 1 \right\} > 0$$

if $\bar{\eta}^* > 0$ and that

$$\frac{\partial \tilde{a}\left(\mu_2, \bar{\eta}^*\right)}{\partial \bar{\eta}^*} = \frac{\tilde{a}^{-\varphi}}{1-\varphi} \mu_2 \varepsilon^l \left(\frac{a^l}{\varepsilon^l}\right)^{1-\varphi} \left[\left(1+\bar{\eta}^*\right)^{1-\varphi} - 1\right] G'\left(\bar{\eta}^*\right) > 0$$

if $\bar{\eta}^* > 0$ and $G'(\bar{\eta}^*) > 0$.

Using the definitions of $\tilde{a}(\cdot)$ and $e(\cdot)$, define the function $a(\cdot)$, the marginal cost in terms of labor of producing one unit of the aggregate consumption, C, given that the current and future states are (μ_2', μ_2). Then function a is given by

$$a\left(\mu_2, \mu_2'\right) = \tilde{a}\left(\mu_2, e\left(\mu_2, \mu'\right)\right).$$

It is immediate from the definition of $a(\cdot)$ that the reduced form planning problem is equivalent to the planning problem.

Using the definitions of a, $\tilde{a}$, and e,

$$\frac{\partial a\left(\mu_2, \mu_2'\right)}{\partial \mu_2} = \frac{\partial \tilde{a}\left(\mu_2, e\left(\mu_2, \mu_2'\right)\right)}{\partial \mu_2} + \frac{\partial \tilde{a}\left(\mu_2, e\left(\mu_2, \mu_2'\right)\right)}{\partial \bar{\eta}^*} \frac{\partial e}{\partial \mu_2}\left(\mu_2, \mu_2'\right) > 0,$$

$$\frac{\partial a\left(\mu_2, \mu_2'\right)}{\partial \mu_2'} = \frac{\partial \tilde{a}\left(\mu_2, e\left(\mu_2, \mu_2'\right)\right)}{\partial \bar{\eta}^*} \frac{\partial e}{\partial \mu_2'}\left(\mu_2, \mu_2'\right) < 0,$$

which is what we assume in the reduced form planning problem.

The reduced form planning problem makes it easy to solve for aggregate consumption and labor, conditional on the current state (μ_2, ε_2) and a choice of μ_2'. These conditional optimal policies are

$$C = \left(\frac{\varepsilon_2}{Aa\left(\mu_2, u_2'\right)}\right)^{1/\gamma},$$

$$n = a\left(\mu_2, u_2'\right)^{1-\frac{1}{\gamma}} \left(\frac{\varepsilon_2}{A}\right)^{1/\gamma} + \mu_2 \phi.$$

These choices for C and n can be replaced into the period utility of the reduced form planning problem so that the only choice is μ_2'. In the special case of log preferences ($\gamma = 1$) this gives

$$V(\mu_2,\varepsilon_2) = \max_{\mu_2'} \left\{ \begin{array}{l} \log \frac{\varepsilon_2}{A} - \log\left(a(\mu_2,\mu_2')\right) - \frac{A}{\varepsilon_2}\mu_2\phi \\ +\beta E\left[V(\mu_2',\varepsilon_2')\middle|\varepsilon_2\right] \end{array} \right\}.$$

I turn to the analysis of the effect of the "demand shock," ε_2, into reorganization. In the log case, the period return function is additively separable in ε_2 and μ_2'. Thus if ε_2 is i.i.d., then $g(\mu_2,\varepsilon_2)$, the optimal decision for μ_2', is independent of ε_2 and hence the amount of reorganization and μ_2 will be constant. This is to be compared with the partial equilibrium analysis wherein the i.i.d. case reorganization is countercyclical.

Assuming that $-\log(a(\mu_2,\mu_2'))$ is concave in (μ_2,μ_2'), the first-order conditions and the envelope displayed below are sufficient for an interior solution:

$$\frac{1}{a\left(\mu_2,g(\mu_2,\varepsilon_2)\right)}\frac{\partial a}{\partial \mu_2'}\left(\mu_2,g(\mu_2,\varepsilon_2)\right) = \beta E\left[\frac{\partial}{\partial \mu_2'}V\left(g(\mu_2,\varepsilon_2),\varepsilon_2'\right)\middle|\varepsilon_2\right],$$

$$\frac{\partial}{\partial \mu_2}V(\mu_2,\varepsilon_2) = -\frac{1}{a\left(\mu_2,g(\mu_2,\varepsilon_2)\right)}\frac{\partial a}{\partial \mu_2}\left(\mu_2,g(\mu_2,\varepsilon_2)\right) - \frac{A}{\varepsilon_2}\phi.$$

If the "demand shock," ε_2, is persistent in the sense that the distribution of ε_2' is stochastically higher if ε_2 is larger, one can show that $\partial V/\partial\mu_2$ is increasing in ε_2. This, in turn, implies that $g(\mu_2,\varepsilon_2)$ is increasing in ε_2. Thus, in the case where ε_2 is persistent (provided $-\log a$ is concave), reorganization is countercyclical. I have solved this model numerically by making discrete the state space, so that the numerical solution does not depend on the assumption that $-\log a$ is concave. In my numerical examples, I have confirmed that g is increasing in ε_2 for $\gamma \le 1$.

REMARKS ON THE GE MODEL

I have analyzed a planning problem instead of analyzing the decentralized equilibrium of the model. The partial equilibrium KR model is one of monopolistic competition. Given the Dixit-Stiglitz specification that I am using, it is easy to formulate the equilibrium problem corresponding to the economy whose optimal allocation I have analyzed. As is well known, the equilibrium allocation differs from the solution of the planning problem due to the effect of the mark-ups of firms on prices, which end up reducing real wages and output. One simple way to reconcile the decentralized economy and the planning problem is to consider a market economy with lump-sum taxes used to finance a subsidy to employment of $\varphi/(\varphi-1)$, which undoes the effect of the mark-ups. Alternatively, I conjecture that the solution of a modified social problem gives the allocations of the decentralized equilibrium. The modification consists of changing the parameter A in some of the expressions for the return function of the planning problem.

I believe that, given the question at hand, it is advantageous to consider a GE model. Nevertheless, the simple model I consider here is deficient as a model of business cycles. As the reduced form social planning problem makes clear, it is equivalent to a static model with productivity $1/a$, except that productivity is endogenous. For the case of log preferences, the model where a is exogenous and random, employment, n, does not depend on productivity. To see this, notice that the conditional optimal decision rule is given by $n = A/\varepsilon_2 + \phi\mu_2$, so it does not depend on productivity $1/a$ directly. Also, the static model has large variations in interest rates. Introducing capital, as in the standard neoclassical growth model, will solve most of these shortcomings.

REFERENCES

Bernanke, Ben S. "The Jobless Recovery." Remarks at the Global Economic and Investment Outlook Conference, Carnegie Mellon University, Pittsburgh, Pennsylvania, November 6, 2003; www.federalreserve.gov/boarddocs/speeches/.

Koenders, Kathryn and Rogerson, Richard. "Organizational Dynamics Over the Business Cycle: A View on Jobless Recoveries." Federal Reserve Bank of St. Louis *Review*, July/August 2005, *87*(4), pp. 555-79.

Federal Reserve Bank of St. Louis
P.O. Box 442
St. Louis, MO 63166-0442