Separating the Business Cycle from Other Economic Fluctuations

Robert E. Hall

Introduction

The U.S. economy rarely grows smoothly. Periods of variable but positive growth are interrupted occasionally by contractions, usually brief. Macroeconomics is making some progress in understanding the patterns of aggregate fluctuations. The answers seem to be anything but simple. The traditional notion no longer holds that the economy moves along a smooth growth trend with temporary cyclical departures. This notion is badly incomplete as a description of the data. Key variables—real GDP, unemployment, and real returns—display important movements at frequencies below the business cycle but above long-term trend. Modern macro theory helps explain these movements in a view that integrates them with cyclical and trend movements.

Over the past half-century, macroeconomists have found the neoclassical model of the economy useful. That model follows the principles of Robert Solow's growth model together with standard ideas about clearing markets. The neoclassical model lacks a business cycle because it lacks the frictions and information limitations that modern macroeconomics finds helpful in explaining transitory movements of employment and output. In a view that is widespread

today, the neoclassical model governs the evolution of the economy in all but the short run—it provides a baseline for understanding macroeconomic fluctuations.

The view that the business cycle is an add-on to a smoothly evolving neoclassical economy has come under a double challenge. One challenge is factual. The decomposition of variables into a trend and a cycle reveals important components that are neither trend nor cycle. The other challenge comes from models. A macro model describes the effects of exogenous driving forces on the key endogenous variables. The models under most active debate today do not support the trend cycle view.

One branch of the new learning emphasizes the volatility that a strictly neoclassical model predicts, given the magnitudes of the shocks that hit the economy. The enduring contribution of Finn Kydland and Edward Prescott's real business cycle (RBC) model was to make this point. The RBC model emphasized the irregularity of productivity growth, but movements in exogenous spending appear to be roughly equal contributors to the volatility of real GDP. Much of that volatility, in fact, appears to come from the neoclassical mechanisms embodied in the RBC model and not from the transitory departures from neoclassical equilibrium that might be called the business cycle.

Another branch, particularly active in the past few years, builds a coherent view of business cycle facts that elude neoclassical explanation. This work focuses on the labor market, though some of the new ideas also might help explain features of other markets. Its primary objective is an understanding of unemployment, a concept not even considered in the RBC model. The new work can be embodied in dynamic general equilibrium models to develop hybrids, where volatility comes from movements of productivity, exogenous spending, and other driving forces, amplified by variations in unemployment. This branch of macro thinking disclaims the notion that the business cycle is an add-on to the neoclassical model. Models instead deliver unified explanations of fluctuations at all frequencies.

The traditional idea is that neoclassical constructs—production functions, consumption demand functions, labor supply functions, embedded in markets that clear—describe the actual operations of the economy in the longer run. There is an economy that generates variables such as y^*, called potential GDP; u^*, called the natural unemployment rate; r^*, called the natural real interest rate; and so on. Milton Friedman's famous presidential address (Friedman 1968) developed this idea informally. Robert Lucas (1972) formalized the idea in a particular way, in which the *-economy is one with full information, and deviations occur because of imperfect knowledge of the current state of the economy.

Deviations of the actual economy around the *-economy constitute the business cycle. They arise from some kind of disequilibrium relative to the neoclassical equilibrium of the *-economy. As Lucas showed, the actual economy is in equilibrium, once limitations on information are included in the model. When I refer to the transitory disequilibrium of the business cycle, I mean specifically in relation to the neoclassical benchmark.

I draw an arbitrary line separating neoclassical principles from the full set of ideas that constitute modern economic science. Much of non-neoclassical economics, as I use the term, involves imperfect information, as in Lucas' model and in the worker-job matching models that have added much to the understanding of unemployment in recent years. Although non-neoclassical economics is sometimes casual in stating its theories, the parts that I consider here are just as rigorous as anything neoclassical.

In the early years of what Paul Samuelson called the "neoclassical synthesis," the *-economy was viewed as generating smooth trends, as described by Solow's growth model, the keystone of neoclassical macroeconomics. Short-run movements around the smooth trend were transitory, the result of imperfect information, delayed adjustment of prices, or other non-neoclassical features of the economy.

An important milestone in the unfolding breakdown of the neoclas-
sical synthesis was Kydland and Prescott's (1982) discovery that the
*-economy is anything but smooth, once the actual volatility of
productivity growth is included in the model. Their real business cycle
model generates substantial short-run volatility from relentlessly
neoclassical principles. This discovery forbids extracting the disequilib-
rium cyclical movements as deviations from a smooth trend. Instead,
one would have to solve the *-model and calculate deviations from the
volatile *-variables. I'm not sure that this lesson fully has informed the
community of practical macroeconomists who try to use signal-extrac-
tion methods based on statistical characterizations of the *-variables as
moving smoothly over time.

The second element in the breakdown is the high persistence of the
deviations of actual from neoclassical performance. The puzzle is most
visible in unemployment, a distinctly non-neoclassical variable.
Unemployment has large low-frequency swings—low in the 1950s
and 1960s, high in the 1970s and 1980s, low again in the 1990s and
2000s. The idea is unpalatable that these movements are the result of
transitory cyclical forces, for they are only barely transitory. The stan-
dard view that the *-economy explains longer-run movements seems
to call for adding low-frequency movements of unemployment to the
*-economy—that is, to create a model of the natural rate of unemploy-
ment that permits slow-moving changes.

Unemployment is not a feature of any neoclassical economy. Expla-
nations of even a constant natural rate of unemployment invoke
distinctly non-neoclassical mechanisms of matching friction, the result
of imperfect information among workers about available jobs and
among employers about available workers. Theories of matching and
unemployment do not deliver a distinction between transitory and
slow-moving elements. They do not contain an object called the
natural rate. They are unitary theories of unemployment. A careful look
at detailed historical data finds no discrepancies—no special factors in
the 1970s and 1980s that would boost unemployment. Rather,
turnover in the labor market was at normal levels in the 1970s and

1980s, and unemployment was high because jobs were harder to find. The period resembled a long recession. The mechanism at work in a recession is close to permanent, not transitory.

I conclude that neoclassical principles properly applied in an environment with volatile driving forces deliver predictions rather different from the smooth growth implicit in most thinking, based on the cycle-over-trend model. Sudden movements of GDP and other variables are not necessarily part of the disequilibrium business cycle—they may reflect the neoclassical response to shifts in productivity and exogenous spending. And non-neoclassical forces in the labor market may result in long-lasting, smooth changes in unemployment that are not distinguishable from the more rapid movements observers earlier have assigned to transitory disequilibrium.

Building new models of the labor market that come to grips with the observed level of volatility of employment and unemployment is an active current area of research. Subtle changes in the economic environment, such as changes in the distribution of information known to one side of the employment bargain but not to the other, can cause large changes in unemployment. And these changes can be long-lasting— they may play an important role in the subcyclical movements of the labor market that are so prominent in the data but escape explanation in existing models.

Although research on the puzzles of financial returns is even more active than in labor market volatility, I am not confident that we are making as much progress. Even the simplest observations, such as the equity premium, defy explanation even from new ideas. And the volatility of the stock market is far above the prediction of any sensible general equilibrium macrofinance model.

The conduct of monetary policy is generally outside the scope of this paper. Some of the lessons are worth mentioning briefly. The traditional view encouraged policymakers to think that the *-economy delivers normal or neutral values of real variables. The idea was that

these reveal the stance of the economy and of policy. If unemployment exceeds the natural rate, real GDP falls short of potential. And if the real interest rate is below its natural level, then the economy is in a cyclical slump and the stimulus of the low interest rate is appropriate. As unemployment and output return to their * levels, monetary policy should move into neutral and the interest rate should be moved to its neutral, * level, according to the neoclassical synthesis.

The new view makes real variables less useful in the formulation of monetary policy. Only an elaborate, realistic version of the RBC model can deliver values of y^* and r^* that take proper account of the movements of productivity and exogenous spending. Even that model does not know how to deal with movements of unemployment. New thinking about unemployment is far too primitive to help in that area, and, in any case, does not support the whole idea of constructing u - u^*. I believe that the idea of a *-model gradually will disappear from macroeconomics and from policymaking.

The Taylor rule has proven to be an exceptionally useful tool for describing monetary policy. A generic rule might be written:

$$r_t = ar^*_t + b(\pi_t - \pi^*) + c(u_t - u^*) + dr_{t-1}. \tag{1}$$

Here, π_t is the rate of inflation, and π^* is the target rate of inflation. The presence of r^* and u^* (or y^*) in the Taylor rule creates problems, according to the evidence and thinking discussed in this paper. Even in a model where these constructs are meaningful, evaluating them in real time is likely to involve serious errors. And the correct model may not involve *-variables at all. The new view points in the direction of a Taylor rule that adapts the interest rate to inflation (positive b), but does not use a base or neutral value of the real interest rate ($a = 0$) and does not try to respond to cyclical variables, such as the unemployment gap ($c = 0$). Rather than set the interest rate to a measure of the neutral real interest rate, the Taylor rule would raise the nominal rate from its earlier value whenever inflation threatened to exceed its target ($d = 1$).

Sources of volatility

Decomposing the sources of volatility in real GDP

I begin by demonstrating—in a model-free environment—the basic point that neoclassical forces, notably productivity growth, predict a good deal of volatility. The *-economy does not evolve along a smooth trend.

Solow's growth decomposition provides a good framework, demonstrating this point with respect to the sources of movements in real GDP:

$$\Delta Y_t = \Delta A_t + \alpha_t \Delta N_t + (1 - \alpha_t) \Delta K_t. \tag{2}$$

Here, Y_t is real GDP, A_t is an index of productivity, N_t is annual hours of work adjusted for quality, K_t is the flow of capital services, and α_t is labor's share in factor income, taken as an estimate of the elasticity of output with respect to hours of work. Δ is the annual percent rate of growth of the corresponding variable.

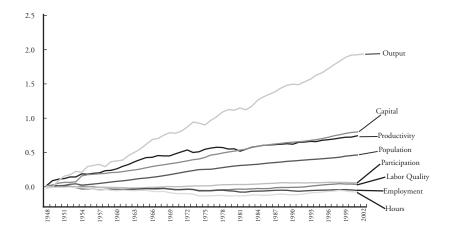Quality-adjusted hours of work further decompose as:

$$\Delta N_t = \Delta L_t + \Delta P_t + \Delta(1 - u_t) + \Delta H_t + \Delta Q_t. \tag{3}$$

Here, L_t is the size of the working age population, P_t is the labor force participation rate, u_t is the unemployment rate, H_t is hours per worker, and Q_t is the quality of the average hour of work, derived from the education mix of workers. Thus, one can examine a seven-factor breakdown of the sources of fluctuations in output:

$$\Delta y_t = \Delta A_t + \alpha_t(\Delta L_t + \Delta P_t + \Delta(1 - u_t) + \Delta H_t + \Delta Q_t) + (1 - \alpha_t)\Delta K_t \tag{4}$$

The Bureau of Labor Statistics prepares data on the total private economy based on a detailed application of Solow's method. Chart 1 shows the growth of the log of output and the seven components for the period since 1948. Chart 2 shows the same data after removing a linear trend from each series.

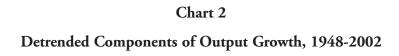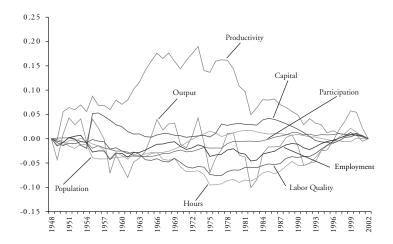**Chart 1**

Components of Output Growth, 1948-2002

One of the most striking features of the data is the large role of productivity in output fluctuations. Chart 2 shows that productivity grows along anything but a smooth path of constant increase. Fluctuations in productivity are important at low, medium, and high frequencies. The variables most closely associated with normal ideas of the business cycle—employment (1 minus the unemployment rate) and hours per worker—are only part of the story of short-run fluctuations. And those two variables have important lower-frequency movements.

I measure the contributions of the variances of annual growth of the various components by their covariances with output growth. This gives an exact additive decomposition of explanatory contributions to the variance of growth. I measure growth over one-, three-, and seven-year periods. Table 1 shows the results.

Almost half of the variance of annual output growth arises from the variance of productivity growth, according to Table 1. This table is biased upward to some extent because the Bureau of Labor Statistics (BLS) measures productivity growth as a residual from measured

Chart 2

Detrended Components of Output Growth, 1948-2002

output growth, so errors in output growth appear in both variables and bias the covariance upward. A second source of bias is the endogenous response of productivity arising from increasing returns or unmeasured changes in capital utilization. Nonetheless, by all appearances, variations in the rate of productivity growth are a major source of variations in GDP growth. Not surprisingly, productivity growth variability accounts for an even larger fraction of the variability of output growth over three- and seven-year periods.

Population, labor-force participation, and labor-quality growth have small variances, so they account for little of the variation in output growth. Variation in growth of the capital stock accounts for a moderate amount of output variability for one- and three-year differences, but essentially none for seven-year differences.

Table 1

Decomposition of Sources of Variation in Output Growth, 1948-2002

Variable	Data Source	Weight	Contribution (Percent)		
			1-Year Changes	3-Year Changes	7-Year Changes
Output growth	Bureau of Labor Statistics, Multifactor Productivity Index, Private Business				
Productivity growth	Bureau of Labor Statistics, Multifactor Productivity Index, Private Business	1	48	52	60
Growth of working-age population	Bureau of Labor Statistics, Multifactor Productivity Index, Private Business	Labor share	-4	-4	4
Growth of labor-force participation rate	Current Population Survey, Labor Force Divided by Civilian Non-Institutional Population	Labor share	0	2	1
Growth of employment rate, 1-unemployment rate	Current Population Survey, Employment Divided by Labor Force	Labor share	18	21	24
Growth of hours per worker	Bureau of Labor Statistics, Multifactor Productivity Index Data, Private Business	Labor share	17	15	13
Growth of quality of the average hour of work	Residual from Multifactor Productivity Index, Private Business, Hours of Work with Quality Adjustment and Bureau of Labor Statistics, Major Sector Productivity Data, Private Business, Hours of Work	Labor share	7	4	-1
Growth of capital services	Bureau of Labor Statistics, Multifactor Productivity Index Data, Private Business	1-labor share	12	10	-2

The two cyclical variables, the employment rate and hours per worker, jointly account for about a third of the variability of annual output growth. As expected for a cyclical variable, variations in hours per worker account for less of the variation in three- and seven-year changes in output than for one-year changes. But the same is not true for the employment rate. Its role in output variability is actually a little higher for the longer differences.

Table 1 little resembles the findings for an economy with smooth underlying growth buffeted by transitory recessions and recoveries. Smooth productivity growth would have little volatility in one-year changes, so it would not account for much of the variability of annual

growth of output. Transitory cyclical variables—the employment rate and hours—would have large roles in annual differences. For seven-year changes, on the other hand, the transitory components would have little role, and variations in productivity growth, though small, would account for all of the similar variations in output growth.

As Chart 2 demonstrates, the primary reason for the findings of Table 1 is that all of the components of output growth are quite persistent. The main difference between productivity and the employment rate, for example, is that productivity is much more volatile. The persistence of fluctuations is comparable between the two variables. Consequently, Table 1 does not assign high explanatory power to variations in the employment rate for one-year differences and high power to variations in productivity growth for long differences.

Volatility in macroeconomic models

Next, I will investigate the potential volatilities of a neoclassical economy model and a model of an economy with the same fundamentals, but with frictions in its labor market. I will compare both volatilities to those of the U.S. economy over the past half-century.

Table 2 shows the standard deviations of key macro variables over the period 1947 through 2004. The variables are chosen to describe a non-stationary economy—one with stochastic variations in its growth rate—but with a stationary unemployment rate and stationary exogenous disturbances to spending. The U.S. economy fits this description. The first two rows describe two major driving forces, productivity and exogenous spending. The latter is the sum of government purchases of goods and services and net exports, stated as a ratio to consumption. Both exogenous spending and consumption are nonstationary, but the ratio is stationary, though quite persistent. I take net exports as exogenous, though, of course, in the global view they are endogenous. Much of the volatility of exogenous spending arises from military spending, reasonably taken to be exogenous.

Table 2

Volatility of Macro Variables, 1948-2002

Variable	Data Source	Units	Value
Productivity growth	Bureau of Labor Statistics, Multifactor Productivity Index, Private Business	Quarterly standard deviation, percent, inferred as half the annual standard deviation	0.94
Government purchases plus net exports as a ratio to consumption	Bureau of Economic Analysis, National Income and Product Accounts, Table 1.1.5. Nominal Gross Domestic Product	Standard deviation of quarterly data, percent	5.01
GDP growth	Bureau of Economic Analysis, National Income and Product Accounts, Table 1.1.3. Real Gross Domestic Product, Quantity Indexes	Standard deviation of quarterly data, percent	1.00
Consumption growth	Bureau of Economic Analysis, National Income and Product Accounts, Table 1.1.3. Real Gross Domestic Product, Quantity Indexes	Standard deviation of quarterly data, percent	0.85
Investment/capital ratio	Growth Ratio of the Real Capital Stock from Bureau of Economic Analysis, Fixed Asset Tables, Table 1.2. Chain-Type Quantity Indexes for Net Stock of Fixed Assets and Consumer Durable Goods	Standard deviation of annual data, divided by 4, percent	0.22
Unemployment	Bureau of Labor Statistics, Unemployment Rate, 16 years and older	Standard deviation of quarterly data, percent	1.53

Kydland and Prescott (1982) revolutionized thinking about the behavior of the aggregate neoclassical model. Prior to their real business cycle or RBC model, the neoclassical model took the form of a growth model, with smooth exponential growth of population and productivity. Kydland and Prescott found that fluctuations in productivity of realistic magnitude induced realistic movements of some macro variables, including real GDP. As I will demonstrate shortly, this finding holds for conventional values of parameters, including the elasticity of labor supply. Kydland and Prescott found that explaining the observed volatility of employment required highly elastic labor supply, a controversial property. Their need to assert unrealistically elastic labor supply derives from their neglect of unemployment, I believe.

Hall (2005d) describes an aggregate modeling framework suited to the investigation of volatility. Here, I start with a version of the model that is strictly neoclassical. The model resembles Kydland and Prescott's in most respects, except that I use a standard specification for capital adjustment costs in place of their specification with time to build (this difference has almost no effect on the results). The most important difference from Kydland and Prescott's work is that labor supply is inelastic. I calculate the exact equilibrium of the stochastic model, rather than using a log-linearization. For many purposes, a log-linear approximation is perfectly satisfactory, but not for the measurement of volatility.

In the model, consumers live forever and make consumption plans that maximize expected utility. Utility each period is a constant elastic function of consumption. The inter-temporal elasticity of substitution is 0.4, and the coefficient of relative risk aversion is 2.5. The technology is Cobb-Douglas. Each person works the same amount each quarter, independent of the wage or the person's wealth. Capital accumulation incurs quadratic adjustment costs such that the quarterly growth rate of capital is one-third of the deviation of Tobin's q from one. All of these parameter values are standard.

A productivity shock with a mean of 0.33 percent and a quarterly standard deviation of 0.94 percent disturbs the economy each quarter. The shocks are cumulative, so that the level of productivity is a random walk with upward drift. An exogenous spending shock with a mean of 32 percent of consumption and a standard deviation of 5 percent of consumption also disturbs the economy. The shock is highly persistent but eventually dies away. The means, standard deviations, and persistence of the two shocks match the data underlying Table 2.

Table 3 shows the volatilities of the key variables. The first column repeats the measures from Table 2. The second column shows the computed volatilities from the neoclassical model. The model falls a little short in matching the volatility of output. It overstates the volatility of consumption by about the same amount. It also overstates the

Table 3

**Volatilities of Key Macro Variables, Two Models
Compared to Actual**

	Actual	Neoclassical Model	Model with Labor-Market Friction
	Standard Deviations		
Output growth	1.00	0.84	0.83
Consumption growth	0.85	1.06	1.03
Investment/capital ratio	0.22	0.26	0.25
Unemployment	1.53	0.00	1.55

volatility of investment somewhat. None of these failings is very serious. With plausible modifications, the neoclassical model could match the volatility of output, consumption, and investment essentially perfectly. The neoclassical model matches those volatilities reasonably well without bringing in the mechanism normally thought to be at the center of the business cycle. Employment remains a constant fraction of the population, and unemployment remains at zero at all times in the model.

The right-hand column reports on the properties of a non-neoclassical model that shares many of the same elements and parameter values of the neoclassical model, but it includes a specification of the labor market with endogenous and quite variable unemployment. The production function and preferences in the model are the same as before, but the labor market has important limitations on information. I will spell out more of the details of the model in a later section. The calibration of the model places the average unemployment rate at its historical level of 5.5 percent and also matches the observed volatility of about 1.5 percentage points.

The key point of Table 3 is that adding a realistic specification of the labor market does not make much difference for the implied volatility of GDP, consumption, or investment, even though it helps

a lot with respect to the volatility of unemployment. Adding variable employment actually reduces the volatility of output growth by a tiny bit, but this finding should not be taken seriously because it rests on the extreme assumption of zero elasticity of labor supply. Rather, the conclusion is that employment variation is likely to be a small part of the total picture of the volatility of output growth. There is much more to the story of output growth than the business cycle captured by unemployment fluctuations.

The finding that much of the volatility of the economy arises from neoclassical sources greatly complicates the extraction of the transitory cyclical components of the movements of macro variables. The *-economy captured by the middle column of Table 3 is hardly a smooth trend. Most of the action in the economy arises from sources other than the business cycle.
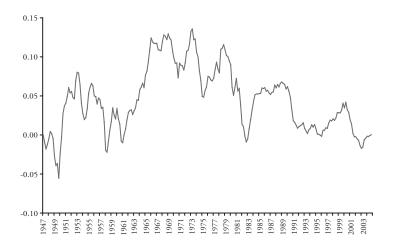
Properties of key macro variables

This section takes a closer look at three macro variables that play key roles in discussions of monetary policy, especially in connection with the Taylor rule. Some forms of the rule call for the central bank to set an interest rate based on a measure of the position of the economy in the business cycle—the gap between real GDP and potential or the deviation of unemployment from the natural rate—and also based on a measure of the neutral level of the real interest rate. Studying the histories of the variables illustrates the challenges in providing these measures.

Real GDP

Chart 3 shows real GDP in terms of the deviation of its log from a linear trend. I remove the trend to bring out the cyclical and subcyclical movements. The business cycle is readily apparent in Chart 10. In addition, real GDP has large subcyclical movements. The normal or neutral level of real GDP usually is called "potential" GDP. Many discussions relate potential GDP to the level predicted by a

Chart 3

Deviations of Log Real GDP Around a Linear Trend

neoclassical *-model. In practice, however, "potential GDP" often is taken as the smooth component of a two-component statistical decomposition, where the second component is the business cycle. The Hodrick-Prescott filter is spectacularly successful in separating an historical series for real GDP into convincing slow-moving trend and higher-frequency cycle components.

Orphanides and van Norden (2002) investigate smoothing methods, including Hodrick-Prescott, for recovering potential GDP in real time, as required for use in a Taylor rule. They demonstrate the difficulty of extracting the current value of the smooth trend from just past data. The great success of Hodrick-Prescott comes in part from its use of future data—it is a two-sided filter. Data revisions are also a major issue for calculating output gaps in real time.

The results in the second section of the paper cast serious doubt on the smoothing or signal-extraction approach to determining potential GDP and thus the GDP gap. The neoclassical model does not predict that GDP will evolve smoothly. Instead, one would need to solve the neoclassical model to obtain Y_t^*, a volatile series. Notice that Kydland
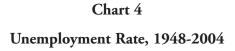
and Prescott studied data after applying the Hodrick-Prescott filter. They interpreted the higher-frequency movements of output remaining after filtering as the equilibrium of their neoclassical RBC model, not deviations from a neoclassical equilibrium. Even macroeconomists like me, who find room for improvement in the RBC model's treatment of the labor market, find persuasive the RBC model's showing that a neoclassical model predicts substantial volatility of output.
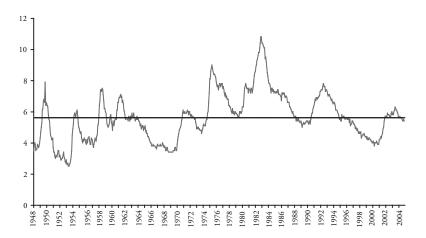
My conclusion is that a central bank should not try to compute a GDP gap in real time and adjust monetary policy according to the current gap. Smoothing approaches to defining potential GDP fail to consider the strong evidence that potential GDP—if there is such a thing—is volatile, not smooth. Macroeconomics is far from having delivered a workable, real-time measurement process for potential GDP.

Unemployment

Chart 4 shows the unemployment rate over the entire period that it has been measured by the current method based on a household survey. The business cycle is a conspicuous feature of the movements of unemployment—each recession results in a spike of unemployment. The series has no trend over nearly 60 years. But unemployment has substantial movements at frequencies below the business cycle. It was high around 1960, low around 1968, high again around 1980, and recently low again.

The chart makes it clear that something more than a simple transitory business cycle drives unemployment. Persistent influences are at work as well. One of them is demographics—the baby boom generation caused a bulge of unemployment as it went through its high unemployment years in the late 1970s. But Chart 5 shows that other highly persistent forces are at work as well. It adjusts unemployment for the age composition of the labor force by taking a fixed-weight index of age-specific unemployment rates. The weights reflect the composition of the labor force in the middle year of the data, 1975. Even with the demographic adjustment, unemployment was high for an extended period in the 1970s and 1980s.
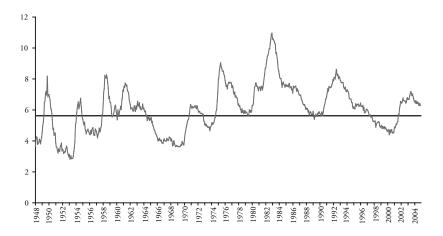
Chart 4

Unemployment Rate, 1948-2004

I ultimately will argue against the two-component view of unemployment, where it is the sum of a slow-moving natural rate and a cyclical component. As usual, the Hodrick-Prescott filter does a marvelous job of finding two such components from historical data. Staiger, Stock, and Watson (1997) estimate a two-component model in a more structural framework, where the transitory component obeys the William Phillips curve principle that inflation rises when unemployment falls below the natural rate. They warn against relying on their estimates of the natural rate in real time—their method yields estimates of the current natural rate with large standard errors.

The real interest rate

The real interest rate displays similar puzzles to those just documented for output and unemployment. The idea that the real interest rate has a stable value r^* and transitory cyclical fluctuations receives even less support than does the same idea for unemployment. Discussion of the real rate is further complicated by the wide variety of ways to measure the concept. Nominal short interest rates have expected real levels and realized real levels. The market for price-level-protected

Chart 5

Unemployment Rate, Adjusted for Age Composition, 1948-2004

Treasury debt gives direct measures of real rates. Longer debt has a real yield and real holding period returns. The stock market has short-, medium-, and long-period real returns. The real return to capital can be measured from a residual or from a model of the marginal product of capital. And the marginal rate of substitution of consumers between this year and the next is another measure that should reveal real returns.

The following equation from macrofinance, embodying the consumption capital-asset pricing model, provides a way to unify all the measures just listed and many others as well:

$$r_{i,t} = \frac{\log \frac{c_t + 1}{c_t} - k_i}{\sigma} + \epsilon_{i,t}. \tag{5}$$

The left side is the ratio of the current to the future price of consumption and the right side is the marginal rate of substitution. Specifically, the left side is the real return earned over a specified holding period. In the results I present, the holding period will be a year. The subscript *i* indexes the wide variety of assets available to the household—debt, equity, physical capital, foreign assets, and the like. On the right, the quantity

$$\frac{\log \frac{c_t+1}{c_t}}{\sigma} \qquad (6)$$

describes the marginal rate of substitution, down to a constant. The parameter σ is the intertemporal elasticity of substitution. The parameters k_i describe the differences in risk and, therefore, return across the assets. Their overall level depends on the volatility of consumption and the rate of impatience. The random variable $\epsilon_{i,t}$ is the surprise in the relation, arising from new information that affects the realized return and next year's consumption. See Hansen and Singleton (1983) for a discussion of the underpinnings of this equation.

I will investigate two ideas about real returns that might be useful in determining neutral or normal levels of returns. One is that a given return is reasonably described by a normal level r_i^* plus an expectation error. In this view, the common element in returns suggested by the consumption growth term in equation (5) is taken to be unimportant (σ large). The second idea is that returns do contain a common source of variation over time. Finance models have this implication. Because those models do not necessarily relate the common time element to the rate of growth of consumption, I will consider an approach that does not assume that the element takes that form.

Table 4 presents basic data on a number of returns and on the marginal rate of substitution. It shows the standard deviation of the variable, the mean over about the past 50 years, and the standard error of the estimated mean. All of the returns are for one-year holding periods, and all are measured as the number of units of consumption achieved by giving up one unit of consumption, investing the proceeds in the asset, and then selling the position and converting back to consumption goods a year later. The variables include two returns to capital. The first takes the earnings of capital as the residual, the difference between the corporate sector's revenue and its noncapital costs. I adjust for corporate taxation, for depreciation, and adjustment costs. The details of the construction of this series appear in Hall (2003). The second measures the return from

the marginal product of capital. For a Cobb-Douglas production function, the value of the marginal product is

$$(1 - \alpha)\frac{pY}{K}, \tag{7}$$

where α, as before, is the elasticity of output with respect to labor; p is the price of output; and K is the capital stock. To derive the return to capital from the value of the marginal product, I equate the latter to the rental price of capital and solve for the return.

Table 4 reveals some of the standard puzzles of finance. The standard deviation of the return to the stock market is vastly greater than the standard deviation of the returns to the underlying activities, as measured by either of the measures of the return to capital. The suspicion of excess volatility of the stock market arises. The average return to the stock market much exceeds the return to Treasury bills—the difference is the notorious equity premium.

The standard errors in the rightmost column of Table 4 show that 50 years of data is enough to give reasonably precise measures of the long-run normal value of the real Treasury bill return and the two measures of the return to capital. If it were true that these returns consisted of a constant and a random surprise, the constant would be well-measured and provide a reliable benchmark for a neutral interest rate or return. But, as I show next, that model does not remotely fit the facts.

The standard errors are much larger for the return to the stock market and for the marginal rate of substitution. Neither is particularly informative about the normal level of real returns.

Charts 6 through 9 show data on annual realized returns for the various assets. Chart 10 shows the realized marginal rate of substitution. Chart 6 plots the realized real return to one-year Treasury bills. To the naked eye, the plot suggests that the return has properties in common with the unemployment rate—in addition to transitory

Table 4

Data on Five Variables that Record Real Returns

Variable	Period	Data Source	Standard Deviation	Mean	Standard Error of Mean
Realized return to one-year Treasury bills	1954-2004	Nominal return on one-year Treasury bills adjusted for change in the NIPA price index for consumption	2.4	2.08	0.34
Realized residual return to corporate capital	1947-2001	Hall (2003)	2.8	4.49	0.38
Return to capital inferred from marginal product of capital	1948-2002	Marginal product of capital estimated by equating the Cobb-Douglas marginal product (the output/capital ratio) to the rental price of capital and solving for the return	1.9	8.03	0.26
Realized real return over one-year holding period, S&P 500	1947-2004	From Robert Shiller, *http://www.econ.yale.edu/ ~shiller/data.htm*	15.5	9.51	2.13
Annual marginal rate of substitution	1948-2001	Equation (5) in this paper, using rate of growth of consumption from NIPA; sigma=0.4	7.8	10.39	0.91

movements lasting only a couple of years, the return has changes in level lasting for a decade or two. In particular, returns averaged around zero until the late 1970s, rose to high levels, and then declined gradually down to current levels around zero again. Only a major feat of signal extraction could find the time-varying neutral, natural, or normal level of this return. Chart 7 shows that much the same conclusion follows for the realized return on corporate capital. In addition to occasional sudden movements of several percentage points, the return has low-frequency movements—high in the 1970s, low in the 1980s, and high again since 1995. Chart 8 provides a related measure with its own low-frequency movements, based on the marginal product of capital. Because this measure is not a residual, it has lower volatility. But it plainly is not a constant plus a serially uncorrelated surprise. Again, signal extraction would be required to determine its normal value. Chart 9 shows the realized real return on the S&P 500. In addition to the large annual deviations demonstrated in Table 4, the plot

suggests a low-frequency component, declining during the first half of the period and rising during the second half. Chart 10 shows the realized marginal rate of substitution, the final candidate to provide a benchmark for the normal level of the real return. It has no visible low-frequency movements, but its volatility is far too high to be useful as a benchmark.

Formal statistical hypothesis testing confirms the impressions from the graphs. Table 5 tests a simple property of a time series that comprises a constant and a serially uncorrelated disturbance. The deviations of the series from its calculated mean should change sign about half the time. The distribution of sign changes would be binomial except for the step of subtracting the mean, which has the effect of slightly increasing the frequency of sign changes. I obtained the null distribution by bootstrap. The table reports p values for three sets of hypotheses. The left column in the table considers the simple model that each series is a constant plus an independent disturbance. The entry is the probability that the observed number of sign changes could have arisen from data obeying serial independence (the test is two-tailed, so the probability includes the possibility of the observed amount of positive serial correlation, with few sign changes, and negative serial correlation, with many sign changes). The hypothesis soundly is rejected for Treasury bills and fairly strongly rejected for the residual return to capital. The test casts some suspicion for the stock market return and the marginal rate of substitution.

The middle column of Table 5 considers whether the time-varying element predicted by finance theory accounts for the rejections in the first column. Because all returns contain the same time-varying component, the difference between any pair should be serially uncorrelated. I subtract the Treasury bill rate from each of the other series and apply the same test, based on sign changes. The hypothesis is rejected decisively for the two returns to capital and moderately for the marginal rate of substitution. Similar results have been found in previous work, starting with Hansen and Singleton (1983). The slow-moving components visible in the figures differ across series.

Chart 6

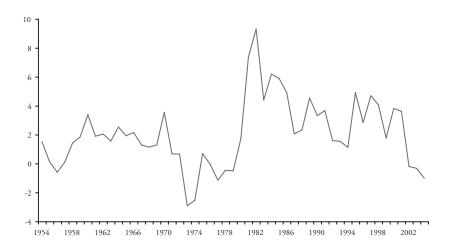## Realized Real Return on One-Year Treasury Bills, 1954-2004

Chart 7

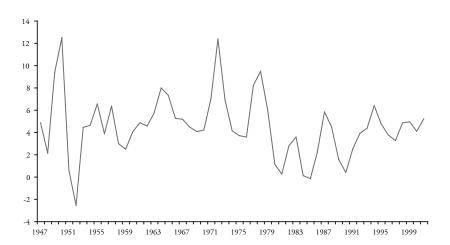## Realized Real Return on Corporate Capital, 1947-2001

Chart 8

Realized Real Return on Capital, Based on Cobb-Douglas Marginal Product of Capital, 1948-2002

Chart 9

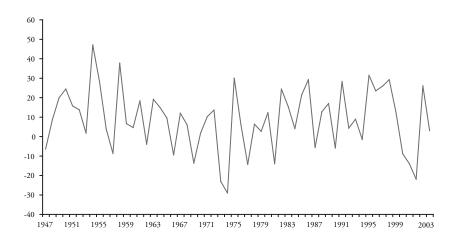## Realized Real Return on the S&P 500, 1947-2004

Chart 10

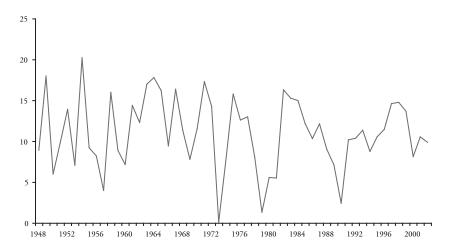## Realized Marginal Rate of Substitution, 1948-2001

Table 5

Tests of the Hypothesis that Returns Have Natural Levels

Variable	Probability that the Series is a Constant Plus Independent Distrubances	Probability that the Difference Between the Treasury Bill Return is a Constant Plus Independent Disturbances	Probability that the Difference Between Series and the Marginal Rate of Substitution is a Constant Plus Independent Disturbances
Realized return to one-year Treasury bills	0.0191		0.2108
Realized residual return to corporate capital	0.0822	0.0003	0.3756
Return to capital inferred from marginal product of capital	0.0000	0.0000	0.2364
Realized real return over one-year holding period, S&P 500	0.3232	0.5124	0.7367
Annual marginal rate of substitution	0.3757	0.2108	

The obstacle to finding a benchmark is more than just accounting for a common slow-moving element.

The right-hand column in the table subtracts the marginal rate of substitution from each of the returns. Because of the high volatility of the marginal rate of substitution, the results are not definitive, but raise the suspicion that the slow-moving component in each return is not the same as the one in the marginal rate of substitution (if there is one in that series).

These results suggest that macro and finance have yet to reach anything like the state of development where a measure of the normal level of any real interest rate is empirically useful or reliable. All measures of the real rate tell different stories.

New thinking about macro fluctuations

A large amount of macro model building continues to superimpose a transitory cycle on a neoclassical *-economy. These models generally assume that sellers—either producers or workers—agree to provide the quantity that buyers choose, given a price or wage that is sticky for a period of time. The *-economy is the same economy without any price-wage stickiness. Because each price or wage adjusts to its *-level after a year or two, the economy tends back to its *-equilibrium. The deviations around that equilibrium are transitory. Woodford (2003) describes the state of this art and relates it in great detail to monetary policy making.

Because this class of models gives the purchaser a call option on the quantity transacted, I call it the "call-option, sticky-price model." Its proponents have put vastly more effort into rationalizing and documenting the stickiness of prices and wages than into rationalizing and documenting the call-option aspect of the relation between buyer and seller. In product markets, one can identify some markets, such as regulated utilities, where the price is set for a period and the buyer chooses the quantity. In retail commerce, merchants set prices and generally fill the demand that materializes, but the period over which prices are

sticky is short in relation to the duration of cyclical fluctuations. In most parts of the labor market, in almost all intermediate product markets, and in some consumer markets, prices are set in idiosyncratic bargains between buyers and sellers, and the buyer has no call option.

The call option, is central to the ability of this class of models to describe the business cycle because it results in volatility of unemployment and other cyclical variables. If workers and employers made contracts that specified both the price and the volume transacted, rather than leaving volume to be determined later, employment and output would not change during the contract. Such an economy would have a puzzle of the excess stability of employment and output, hardly the issue with the U.S. economy.

Barro (1977) made a trenchant criticism of the call-option aspect of the call-option sticky-price model, to which the proponents have not made an effective response, in my view. The call option is not the natural structure for the bilateral relation between buyer and seller. Call options of sufficient duration to match the duration of business cycles do not seem to be widespread. Indeed, I might go farther to say that they hardly seem to exist at all. But it is not my purpose here to criticize the call-option, sticky-price model, but it is, rather, to discuss emerging thinking about other non-neoclassical explanations of macro fluctuations. The new thinking views buyers and sellers as making bilateral bargains. It dispenses with the call option of earlier cycle models, but it may retain sticky prices and wages.

The new thinking concentrates on the labor market, but many of its ideas might apply in product markets, especially intermediate ones. A great deal of progress has occurred in recent years in understanding unemployment, mostly in partial-equilibrium analysis. General equilibrium analysis incorporating the new ideas generally confirms that the insights from partial-equilibrium models carry over.

The equilibria of the new models do not take the form of the equilibrium of the neoclassical model perturbed by a transitory cyclical

component. In this respect they differ fundamentally from the model implicit in Milton Friedman's presidential address, from Lucas' misperceptions model, and from the now-dominant call-option, sticky-price model. The new models do not have a natural unemployment rate, a potential level of GDP, or a normal real interest rate.

The matching model of unemployment

Diamond (1982), Mortensen (1982), and Pissarides (1985) (DMP) launched the view that now accounts for most thinking about unemployment. The model focuses on the employment-unemployment margin and usually takes labor supply in the sense of participation to be inelastic with respect to the payoff from participation. Wage elasticity is easy to add to the model.

In the DMP family of models, the labor market is in continuous equilibrium. No player waits to change a price or allocation once the change is merited. Unemployment arises because job seekers and prospective employers encounter friction that limits their flows of meetings. When a job seeker does meet an employer, the two determine if their prospective relationship has a surplus. The job seeker's reservation wage is the wage of the job foregone by taking the current job. The employer's reservation wage is the net productivity of the job seeker in the employer's job. The surplus is the difference between the employer's reservation wage and the job seeker's, if the difference is positive. Having found a positive surplus, the pair makes a bargain and the job begins. The worker remains in the job until the surplus evaporates. The relation between the worker and the employer is bilaterally efficient.

Employers decide upon a level of recruiting effort—they expend resources whenever the gain from adding another worker exceeds the cost of the effort required to get in touch with the worker. A matching technology relates the job-filling rate to the resources expended by employers. The job-finding rate—the monthly probability that a job seeker will be matched to an employer with a suitable job—moves in

the direction opposite from the job-filling rate. If employers are deploying low levels of effort, jobs are hard to find and the job-finding rate is low. The job-filling rate is correspondingly high, as employers find that candidates for openings are plentiful.

Unemployment depends negatively on the job-finding rate and positively on the job-loss rate. Except for unimportant transition effects immediately after one of the rates changed, two equations determine the unemployment rate:

unemployment entry rate = separation rate x fraction becoming unemployed (8)
    + LF entry rate x fraction becoming unemployed,

$$u = \frac{\text{unemployment entry rate}}{\text{unemployment entry rate + job-finding rate}}. \tag{9}$$

Here the separation rate is the fraction of workers who lose or leave their jobs each period. The LF entry rate is the flow of people into the labor force.

Hall (2005c) reviews a variety of data sources that appear to agree that the separation rate is essentially a constant. Although job losses rise in a recession, the increase is tiny in relation to the high normal levels of separations, which run more than 3 percent per month. The entry rate to unemployment rises in recessions not because of higher separations, but because those separated are more likely to enter unemployment rather than move directly to new jobs, and those entering the labor force are more likely to transit through unemployment.

The job-finding rate and its close cousins, the fractions of separators and entrants who become unemployed, are the key determinants of the unemployment rate. In strong markets, the job-finding rate is high and the fractions of people who become unemployed are low, so unemployment is low.

The labor market is in equilibrium when employers gain the same amount from a new hire as they pay for the recruiting effort that

results in the hire. The market is continuously in this equilibrium. The employer's gain from a match is the productivity of the match less the wage paid to the worker. Thus, the wage is a central determinant of unemployment. If the wage is high, the employer's incentive to create a job and expend efforts to fill it is small. A high wage across the labor market will result in equilibrium with high unemployment. Fluctuations in wages will result in movements in unemployment, and so will fluctuations in productivity if the wage does not change to offset the fluctuations.

What forces determine the wage? At the most fundamental level of economic logic, the wage bargain is indeterminate over a range. Employer and job seeker enjoy a prospective surplus. Any wage between the job seeker's reservation wage (the opportunity cost) and the employer's reservation wage (productivity) will split the surplus and create a bargain that makes both parties better off than they would be if they failed to make the match. A wage-determination mechanism overcomes the indeterminacy by assigning a wage that lies within the bargaining set.

Until quite recently, models in the DMP tradition assumed that the parties split the difference by setting a wage that is a weighted average of the two reservation wages. This wage-determination rule guarantees that the wage lies within the bargaining set. It turns out to make the wage highly flexible. For simplicity, I will assume that the weights are the same and, thus, equal to one-half. When productivity falls, the employer's reservation wage falls by the same amount. On this account, the wage falls by half the amount of the productivity decline. But the job seeker's reservation wage then falls as well because it depends on wages at other jobs. In the new equilibrium, the wage falls by essentially the same as the decline in productivity. The employer's gain remains close to the same, and nothing much happens to the job-finding rate and, thus, the unemployment rate. The model based on the split-the-difference wage rule predicts almost constant unemployment. It leaves all movements in unemployment, those that occur in recessions and those that occur at

lower frequencies, almost completely unexplained. Shimer (2005) made this important finding about the standard DMP model.

Sticky wages in the DMP model

Some of the effort to alter the DMP model to equip it to explain the observed volatility of unemployment has focused on wage determination. Hall (2005b) makes the simple observation that a constant wage satisfies the equilibrium condition, provided that the job seeker's reservation wage remains below the constant wage as the economy's driving forces change and provided that the employer's reservation remains above the constant wage. With a constant wage, unemployment is realistically sensitive to changes in determinants, such as productivity. If productivity falls, employers find the payoff to hiring is lower because the wage they pay is unchanged. They cut back on recruiting activity, the job-finding rate falls, and unemployment rises. A decline in productivity of only a fraction of a percent is enough to raise unemployment by the 2 percentage points typical of a recession.

I call the view that emerges from this line of thought the "equilibrium sticky-wage model." It retains the appealing feature of earlier thinking about sticky wages—the fact that wages in practice seem to have a good deal of inertia. It eliminates the call-option element of earlier sticky-wage modeling, with its property of inefficient disequilibrium. Instead, the new approach describes a full economic equilibrium at all times. No pair of actors faces an opportunity for a Vilfredo Pareto-improving reallocation. No worker loses a job because the wage is too high, in a situation where preserving the match job would make both the worker and the employer better off.

Although the model in Hall (2005b) embodies stickiness of the real wage, the same line of thought rationalizes nominal stickiness as well. A rule or process that governs the evolution of the money wage will result in a full economic equilibrium as long as it delivers a real wage that is within the range—generally fairly wide—between the worker's and employer's reservation wages. With nominal wage stickiness, monetary policy has leverage over the real economy. New models of

the labor market may help understand why this leverage seems to be so strong in the short run and why inflation responds only slowly to monetary control.

The source of unemployment fluctuations in the equilibrium sticky-wage model—and in essentially all of the models currently emerging from the DMP tradition—is quite different from the source in earlier models. In the call-option models, unemployment rises as employers discharge workers whose wages exceed the new values of their marginal products. Job loss is the proximate driving force of rising unemployment in a recession. In the RBC model, employment falls in a recession because incentives have moved against work in the market and toward time spent outside the market. The DMP models bring in a third key determinant—the incentives for job creation.

In the DMP class of models, employers deploy resources to attract workers. When the payoff to employers is high—because the value of the marginal product of labor exceeds the wage by a fair margin—recruiting is enthusiastic and the labor market is tight. Job seekers find new jobs quickly and unemployment is low. In frictional markets, such a gap always exists—it is the employer's share of the search capital that rewards the eventual match. In the DMP models, the amount of the search capital and its sharing between employer and worker are central to the determination of unemployment. Although the canonical DMP model—Mortensen and Pissarides (1994)—views the employer as the active decisionmaker and denies the job seeker any role in choosing the resources devoted to matching, more recent models in the DMP class take a more symmetric view.

Kennan (2005) develops an alternative theory of sticky wages in the DMP framework. In his model, workers are heterogeneous. The job seeker's productivity is known to the employer but is hidden from the job seeker. The job seeker makes a wage bid, which the employer either accepts or rejects; there is no opportunity for further negotiation. The job seeker is in the same position as a bidder in a first-price, sealed-bid auction. Kennan makes assumptions that cause the job seeker to bid a

wage that is insensitive to current conditions. Thus, he reaches a sticky-wage property as a derived conclusion rather than as a bald assumption. Kennan's model delivers a high sensitivity of unemployment to changes in productivity for the same reasons I just discussed in connection with my model with arbitrarily sticky wages.

In a more general version of the model, the job seeker will make a higher bid when conditions are better, using the general principles of the first-price auction theory. Wages will not be completely sticky. Rather, stickiness will be a feature of certain types of distributions. Kennan's model is one of several new models in the DMP class that make outcomes in the labor market depend on features of the distribution of information among participants in the market. Tawara (2004) is an example.

Hall and Milgrom (2005) reconsider the bargaining setup in the standard DMP model. In that setup, the job seeker and employer take an all-or-nothing view about the bargain. Either they make a deal immediately or they make no deal at all. In the latter case, the worker remains unemployed, and the employer loses the benefit for a period of filling the job. We point out that, in reality, people try again to make a deal because if they walk away, they lose the entire surplus. A basic idea of bargaining theory, introduced formally by John Nash, is that the parties' outside options or threat points determine the outcome of bargaining. The standard DMP model links the outcome of the bargain directly to the unemployment rate and other determinants of the outside options. A more realistic view is that a job seeker or employer dissatisfied with a proposed bargain will delay briefly and make a counterproposal. Bargaining theory in this framework—Binmore, Rubinstein, and Wolinsky (1986)—shifts the emphasis from values relating to the labor market to values and costs of delay.

Our model delivers wage rigidity by disconnecting wage bargaining from conditions in the labor market. Once a qualified worker and an employer have found each other and determined that they have a joint surplus, costs of delay, not outside conditions, determine the bargain they make.

The wage does respond to productivity, but only half as much as in the standard model. The result is a strong response of unemployment to productivity and other driving forces. The wage no longer has a strong equilibrating role. If productivity falls, the part of the surplus accruing to employers falls sharply and they cut back on recruiting efforts. The labor market softens dramatically.

Unemployment volatility with flexible wages

Though sticky wages make unemployment highly responsive to driving forces, other mechanisms may be even more important in understanding unemployment volatility. One key issue from equation (8) is the volatility of the fraction of people separating from jobs who become unemployed. Hall (2005c) shows that about half of departures are directly to new jobs, without intervening unemployment. No theory of labor-market dynamics possibly could be complete without consideration of this key flow. Plainly, many job seekers are recorded as employed, not unemployed.

Nagypál (2004) combines an on-the-job search with a number of other key ingredients to achieve quite substantial volatility of unemployment. In her model, workers have different satisfaction levels with their jobs, hidden from employers. Workers hired from unemployment are less desirable because those who form matches will have a lower average job satisfaction. They are more likely to leave the job soon because they search for better jobs while employed. The final key element is a fixed cost of training a new worker. If the employee quits, the employer is deprived of the value of the training cost. Nagypál suggests that it is plausible that the costs from the higher turnover of workers hired from the unemployed considerably more than offsets the easier recruitment of the unemployed. In Nagypál's calibration, the elasticity of the unemployment rate with respect to productivity is about –5. A decline of productivity of 1 percent raises the unemployment rate by about 0.3 percentage points. Thus, on-the-job search and heterogeneous job satisfaction could play an important role in the overall explanation of unemployment volatility.

Hall (2005a) considers a rather different hidden-information problem in the labor market. A job seeker is either qualified or not qualified for a particular job. She has information about her likelihood of being qualified prior to applying for a job with an employer. That information is hidden from employers until they test and otherwise evaluate a job applicant. Making an application is costly to the applicant. Job seekers set a cutoff level of the likelihood and apply for every job that meets the cutoff. Employers know the fraction of applicants who are qualified and expand job openings up to the point that the surplus they enjoy from testing and hiring the average applicant exhausts the testing cost. Job seekers are in equilibrium when the anticipated share of the surplus exhausts the application cost. Once an applicant is tested and found qualified, the job seeker and employer make the standard Nash bargain.

The key determinant of equilibrium in the labor market in the model is the cutoff level of the qualification likelihood. The equilibrium is fragile because a higher cutoff is beneficial to both job seekers and employers. The equilibrium is at the intersection of two curves in surplus-cutoff value space, and the two curves may have almost the same slope. If the cutoff level is low, the market is in an undesirable equilibrium—employers are receiving large numbers of applications from unqualified workers. Employers recruit correspondingly less, so the market is slack. In a slack market, job seekers set low cutoffs because jobs are hard to find. When the cutoff level is high, the market equilibrium induces efficient self-selection. Employers hire enthusiastically because each costly test is likely to yield a new employee who is qualified. Workers set high cutoffs because jobs are easy to find.

This description suggests that the equilibrium is indeterminate, which is definitely a possibility and is not a borderline case. If the equilibrium is determinate and satisfies a standard stability condition, the equilibrium is fragile—it responds sensitively to driving forces.

The driving forces that alter the cutoff qualification level and thus the job-finding rate do not include productivity. Shifts in productivity alter the employer's and worker's surplus in proportion, so the intersection in cutoff-surplus space occurs at the same cutoff level. The most interesting potential driving force is a property of the probability distribution of the signal that job seekers receive about the likelihood of qualification for a job. The property is the relation between the cutoff level adopted by the job seeker and the average likelihood of qualification of applicants employing the rule of applying for every job where the information conveys a likelihood of at least as high as the cutoff. The latter controls the employer's payoff from testing. The elasticity of the ratio of the two is key. If the elasticity is one, equilibrium is indeterminate. Small changes in the elasticity are a potent driving force for large fluctuations in the job-finding rate and other aspects of the labor market.

Sources of recessions

When asked to describe a particular recession or recessions in general, the practical macroeconomist will tell a story that focuses on the collapse of purchases of certain categories of products-producer and consumer durables. For example, all practical accounts of the 2001 recession emphasize the huge decline in high-tech investment. In earlier recessions, declines in home-building were prominent features. On the other hand, more theoretically inclined macroeconomists tend to take a decline in productivity—or at least a pause in the normal growth of productivity—as the central driving force. New ideas discussed here may help bridge this important gap between practical and theoretical macroeconomists.

Some of the new models emphasize the shapes of distributions of hidden information. In the self-selection and on-the-job-search models, employers make decisions knowing the shape of a distribution of job seeker characteristics, but not the hidden value for a particular job seeker. In Kennan's model, a job seeker makes a wage demand without knowing the employer's reservation wage, based on hidden match productivity.

This type of model opens the possibility of subtle driving forces involving changes in the shapes of those distributions.

Changes in distributions may provide the needed link between the practical macroeconomist's notion of a recession and the class of theories considered in this paper. The data show unambiguously that construction, durables, and nondurables manufacturing suffer large employment reductions in recessions, while other industries shrink only slightly or continue to grow—see Hall (2005c).

The older models in the DMP tradition do not provide an immediate analysis linking changes in the industry composition of employment to the aggregate unemployment rate. The rise in unemployment in a recession is the result of diminished job creation among employers in general. If recessions were periods of generally higher reallocation of labor, the traditional DMP model would predict rising recruiting effort in recessions, as employers in other industries took advantage of the availability of workers released from shrinking industries. But the data show amazing declines in recruiting effort in recessions—help-wanted advertising usually falls in half from peak to trough.

The new additions to the DMP class of models may offer some hope of connecting declines in manufacturing and construction to the dramatic rise in unemployment that accompanies every recession. For example, the events leading to a large decline in employment in durables might shift the economy from the favorable equilibrium described in the self-selection model to the unfavorable one. In the favorable equilibrium, the applicants for a job opening are largely people who know they are qualified. Employers waste few resources screening out unsuitable applicants. They are correspondingly enthusiastic about creating jobs, so the market is tight. A subtle change in the distribution of the signal that workers receive about their likelihood of qualification can move the equilibrium perversely. Applicants, finding it difficult to locate any job, apply for jobs where they are less likely to be qualified. Employers are overwhelmed by

applicants and dissipate resources, screening out the unqualified ones. The market becomes slack, with high unemployment.

In Kennan's model, the shape of the distribution of match productivity, a variable observed only by the employer, has two key roles. Job seekers know the distribution but not the realization, so they solve a wage-bidding problem defined by the distribution. Firms earn an informational rent on the difference between the productivity realization and the wage bid. Shifts in the distribution induced by changes in the composition of employment might result in changes in the rent.

The new research offers some new mechanisms that may advance understanding of both the business cycle and the important sub-cyclical movements of unemployment and other key variables. Notice that none of the driving forces of unemployment fluctuations discussed here are explicitly transitory. They do not disappear after firms and workers get around to changing prices and wages. They can explain the highly persistent subcyclical movements that elude the call-option, sticky-price model and the neoclassical model.

My purpose here is to advocate the new models as useful additions to our toolbox of macro fluctuation studies. I do not suggest that we can dispense with the call-option, sticky-price model. In addition, all macroeconomists should be aware of the importance of fluctuations in productivity and exogenous spending, operating through neoclassical channels, in fluctuations at every frequency.

*–Variables in future models

Although the idea of *-variables as the equilibrium of a friction-free economy does not seem to be useful, either in theory or in practice, something like *-variables may be implicit in future macro models. Dynamic systems often have representations in terms of adjustment processes in which current values of variables adjust toward *-values. My remarks in this paper are not intended as a criticism of that type of representation. Rather, I believe that these representations depend critically on the assumptions of the underlying model. For example,

the assumption that sellers grant call options to buyers is central to the standard sticky-price model in use today. New models based on different and perhaps more realistic assumptions have different representations. Further, and most important, all models imply that the *-variables are in constant motion and extraordinarily difficult to infer in real time.

Modeling real returns

Research has been active in developing models that attempt to match the puzzling facts about real returns. The failures of the standard model of the joint behavior of real returns and consumption growth are well-known.

Much of the research has focused on basic aspects of the failure of the standard model. These are the high average return earned in the stock market and the low average return for Treasury bills. The standard model sees both these investments as having low financial risk—their payoffs are not systematically low during bad times, as signaled by low consumption. The reason is that consumption is not terribly volatile. It grows a bit in most years and shrinks only in the rare severe recession. Consumption has not declined, year over year, in the past 25 years and declined only twice, in 1973-1974 and 1979-1980, since 1948. Research that measures risk by consumption growth in the past 50 years finds little demand for insurance against risk and little aversion to risky assets, such as stocks, because bad times have not occurred. It is a paradox that stocks have returned so much more over that period than Treasury bills because investors should not care about the variability of returns to stocks.

One idea proposed by many authors over the years is that investors apply a subjective probability distribution about future consumption that puts some weight on true disasters, even though none has occurred in our lifetime. A simple version is that people know that consumption fell by 20 percent from 1929 to 1933. Even after 50 years of smooth sailing, investors might put some probability on a

repetition of the disaster and know as well that, far from providing insurance against the disaster, the stock market yielded negative returns over the same period. But this explains only a small fraction of the gap between stock and Treasury bill returns. Probability distributions that permit consumption to fall to a billionth of its current value are capable of explaining the gap. Whatever event has this effect must leave the federal government able to meet its Treasury bill obligations, or else the equity premium would not prevail.

Though thousands of papers have tried to explain the equity premium and hundreds more appear every year, it remains a puzzle, as far as I am concerned. The answer may lie in nonstandard or behavioral principles, though I am not aware of any convincing research on this point to date.

Other aspects of returns remain equally puzzling. First, the volatility of financial returns is much higher than models predict. One of the important assets in equation (5) is productive capital—plant and equipment. Charts 7 and 8 show that the return to capital, measured either as a residual or as the marginal product of the Cobb-Douglas technology, is quite smooth. I already have noted that consumption growth is smooth. The nonfinancial variables obey the equation quite well. Further, the general equilibrium model I discussed in the second section can match the volatility of the return to capital and consumption growth.

Financial returns are quite another matter. The volatility of the real return to Treasury bills should be unambiguously *lower* than the volatility of capital, according to the standard model. Markets determine the nominal bill rate in advance—the only expectation error arises from small errors in forming expectations in inflation over the coming year. But, in fact, Chart 6 shows that the Treasury bill rate is rather more volatile than either measure of the return to capital. Moreover, while the average level of the return to capital remains roughly constant, the real Treasury bill rate is low for long periods and then high for long

periods. Its low-frequency movements are paradoxical in the light of the standard model. This topic has attracted almost no research.

The volatility of the return to debt is nothing compared to the volatility of the stock market. Shiller (1981) showed that the stock market varies far too much to fit a model in which investors discount future dividends at a constant rate. Though many of us have suspected ever since that this result supports the hypothesis of excess volatility, subsequent commentary made it clear that variations in discount rates need to be brought into the analysis. Working this out takes us back to equation (5). Excess volatility of the stock market shows up as movements of stock market returns that are inconsistent with the restrictions in that equation. Although research has consistently rejected those restrictions, the potential sources of the rejections are numerous. Irrational investors are only one. Variation over time in the k_i parameters is an explanation potentially consistent with standard ideas about rationality.

A totally different line of attack suggests excess volatility of financial markets. By buying a suitable portfolio of equity and debt claims on a corporation, an investor can synthesize ownership of a share of the corporation's real assets. The return to the portfolio should be the same, period by period, as the return to those assets. In fact, the return to the financial portfolio differs wildly from the return to underlying corporate assets, both in its year-to-year movements and in its volatility. The financial claim on the real assets leads a completely separate life from the assets themselves, both in terms of levels of value and in terms of annual returns.

In some ways, the most telling evidence of the failure of our basic ideas about financial markets is the difference between the value of corporations in financial markets and the apparent value of the capital the corporations own. Chart 11, taken from Hall (2001), shows the ratio of the market value of all financial claims (equity and debt outstanding net of equity and debt held) to the estimated value of the business assets of U.S. nonfinancial corporations. This measure is

Tobin's *q*. The ratio has huge subcyclical movements that do not fit any simple theory. Tobin offered the hypothesis that *q* fluctuates because corporations earn rents on capital made temporarily scarce or plentiful by adjustment costs. But the costs implied by the magnitude and persistence in the graph are far beyond the plausible. My work suggested the corporations were accumulating intangibles, which, if added to the denominator of *q*, would result in a measure that stayed closer to its logical value of one. But this view fails to explain the period from the early 1970s to the late 1980s when *q* was so far *below* one.

Concluding remarks

Discussions of practical macroeconomics, especially those relating to monetary policy, rely heavily on three concepts: potential GDP, the natural rate of unemployment, and the normal real interest rate. To measure potential GDP, practitioners use a statistical smoothing approach, such as the Hodrick-Prescott filter, or a neoclassical growth model. The smoothing approach makes little sense in view of the evidence that a fully neoclassical economy, buffeted by changing rates of productivity growth and changing levels of exogenous spending, does not evolve at all smoothly. In fact, the evidence shows that most of the variation in real GDP growth over short and medium horizons arises from productivity growth. The approach based on a neoclassical growth model enters the thicket of controversy over the proper specification of a real business cycle model. Potential GDP is not a useful guide to making monetary policy.

The U.S. unemployment rate is a stationary variable. In that respect, the idea of a natural rate is on firmer ground than is the idea of potential GDP. But the natural rate cannot be a constant, or else one would be required to believe that the economy spends long periods, such as the 1970s and 1980s, with unemployment above the natural rate and other long periods below the natural rate. Because the neoclassical model has zero unemployment, the model provides no guidance about variations in the natural rate. Modern theories of unemployment say nothing about the natural rate either—they

Chart 11

**Ratio of the Market Value of Corporations to the Value of
Corporate Assets (Tobin's *q*)**

regard every observed level of unemployment, whether in a recession,
boom, or normal period, as a full economic equilibrium. These theo-
ries are unitary. They do not separate unemployment into
equilibrium and cyclical components. Again, the natural unemploy-
ment rate is not a useful concept in the execution of monetary policy.

   New thinking about unemployment departs from the framework in
Friedman (1968), in that Friedman distinguished between the natural
rate and the actual rate of unemployment. He viewed the natural rate
as varying over time, based on "the actual structural characteristics of
the labor and commodity markets, including market imperfections,
stochastic variability in demands and supplies, the cost of gathering
information about job vacancies and labor availabilities, the costs of
mobility, and so on" (p. 8). The new theories see those forces as deter-
mining the *actual* unemployment rate. Nothing in new thinking has
shaken Friedman's basic message, however, that unemployment is

invariant to the monetary policy regime. In that sense, natural rate theory is thoroughly embedded in all modern thought.

The situation with respect to the real interest rate is similar. Wicksell's natural or normal interest rate, as distinguished from the actual market rate, is not a feature of modern macrofinance models. Each of the many real interest rates in the economy moves differently—it does not obey even the relatively unrestrictive principles of basic finance models. We are not equipped to judge when monetary policy is neutral in terms of interest rates.

None of these conclusions stands in the way of intelligent monetary policy making. Under Alan Greenspan's stewardship, the United States has achieved remarkably low levels of inflation and inflation volatility, despite the lack of real reference points. We do not need to know the GDP gap, the unemployment gap, or the neutral real interest rate to keep the price level near constancy.

References

Barro, Robert J. "Long-Term Contracting, Sticky Prices, and Monetary Policy," *Journal of Monetary Economics,* 3 (3), July 1977, pp. 305–316.

Binmore, Ken, Ariel Rubinstein, and Asher Wolinsky. "The Nash Bargaining Solution in Economic Modeling," *RAND Journal of Economics*, 17 (2), summer 1986, pp. 176–188.

Diamond, Peter A. "Aggregate Demand Management in Search Equilibrium," *Journal of Political Economy*, 90 (5), October 1982, pp. 881–894.

Friedman, Milton. "The Role of Monetary Policy," *American Economic Review*, 1968, 58 (1), pp. 1–17.

Hall, Robert E. 2001. "The Stock Market and Capital Accumulation," *American Economic Review*, 91, December, pp. 1185–1202.

_____. "Corporate Earnings Track the Competitive Benchmark," December 2003. NBER Working Paper 10150.

_____. 2005. "The Amplification of Unemployment Fluctuations through Self-Selection," NBER Working Paper 11186.

_____. 2005. "Employment Fluctuations with Equilibrium Wage Stickiness," *American Economic Review*, 95 (1), March, pp. 50–65.

_____. 2005. "Job Loss, Job Finding, and Unemployment in the U.S. Economy over the Past Fifty Years," *NBER Macro Annual,* forthcoming.

_____. 2005. "The Labor Market and Macro Volatility: A Nonstationary General Equilibrium Analysis," Hoover Institution, Stanford University. June.

Hall, Robert E., and Paul R. Milgrom. "The Limited Influence of Unemployment on the Wage Bargain," February 2005. NBER Working Paper 11245.

Hansen, Lars Peter, and Kenneth J Singleton. "Stochastic Consumption, Risk Aversion, and the Temporal Behavior of Asset Returns," *Journal of Political Economy*, 91 (2), 1983, pp. 249–65.

Kennan, John. "Private Information, Wage Bargaining, and Employment Fluctuations," Department of Economics, University of Wisconsin, May 2005.

Kydland, Finn, and Edward C. Prescott. "Time to Build and Aggregate Fluctuations," *Econometrica*, 50 (00), November 1982, pp.1345–1370.

Lucas, Robert E. "Expectations and the Neutrality of Money," *Journal of Economic Theory*, 4 (2), April 1972, pp. 103–124.

Mortensen, Dale T. "Property Rights and Efficiency in Mating, Racing, and Related Games," *American Economic Review*, 72 (5), 1982, pp. 968–979.

Mortensen, Dale T. and Christopher Pissarides. "Job Creation and Job Destruction in the Theory of Unemployment," *Review of Economic Studies*, 61 (0), 1994, pp. 397–415.

Nagypál, Éva. "Worker Reallocation over the Business Cycle: The Importance of Job-to-Job Transitions, Part 2: Theory," Department of Economics, Northwestern University, August 2004.

Orphanides, Athanasios, and Simon van Norden. "The Unreliability of Output-Gap Estimates in Real Time," *Review of Economics and Statistics*, 84 (41), pp. 569-583, November 2002.

Pissarides, Christopher. "Short-Run Equilibrium Dynamics of Unemployment, Vacancies, and Real Wages," *American Economic Review*, 75 (4), 1985, pp. 676–690.

Shiller, Robert J. "Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends?" *American Economic Review*, 71 (3), pp. 421-436, June 1981.

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

Staiger, Douglas, James H. Stock, and Mark W. Watson. "How Precise Are Estimates of the Natural Rate of Unemployment," in Christina D. Romer and David H. Romer, eds., *Reducing Inflation: Motivation and Strategy*, Chicago: University of Chicago Press for the National Bureau of Economic Research, 1997, pp. 195–246.

Tawara, Norikazu. "A DMP Model with Asymmetric Information," Departments of Economics, University of Chicago, August 2004.

Woodford, Michael. *Interest and Prices*, Princeton: Princeton University Press, 2003.