The Long Slump

Robert E. Hall*
Hoover Institution and Department of Economics,
Stanford University
National Bureau of Economic Research
rehall@stanford.edu; website: Google "Bob Hall"

December 3, 2010

Abstract

In a market-clearing economy, declines in demand from one sector do not cause large declines in aggregate output because other sectors expand. The key price mediating the response is the interest rate. A decline in the rate stimulates all categories of spending. But in a low-inflation economy, the room for a decline in the rate is small, because of the notorious lower bound of zero. I build a general-equilibrium model that focuses on the behavior of an economy when the nominal interest rate is pinned at zero. Equally important is that the real rate is pinned at a rate above the market-clearing rate because inflation responds only weakly to the presence of slack. I concentrate on three closely related sources of declines in demand: the buildup of excess stocks of housing and consumer durables, the corresponding expansion of debt that financed the buildup, and financial frictions that resulted from the decline in real-estate prices. The model characterizes rationing of customers in the output market when the interest rate is pinned at zero and connects the rationing to the labor market. It provides a coherent rationale for the common-sense notion that the reason that employers don't hire all available workers during a slump is that they don't have enough customers. I demonstrate the empirical relevance of the three driving forces.

*This paper provides background for my AEA presidential address, January 8, 2011, Sheraton Denver Downtown Hotel, 4:40 pm in the Grand Ballroom. I am grateful to Steven Davis, Mark Gertler, Robert Lucas, Richard Rogerson, David Romer, and Robert Shimer for comments.

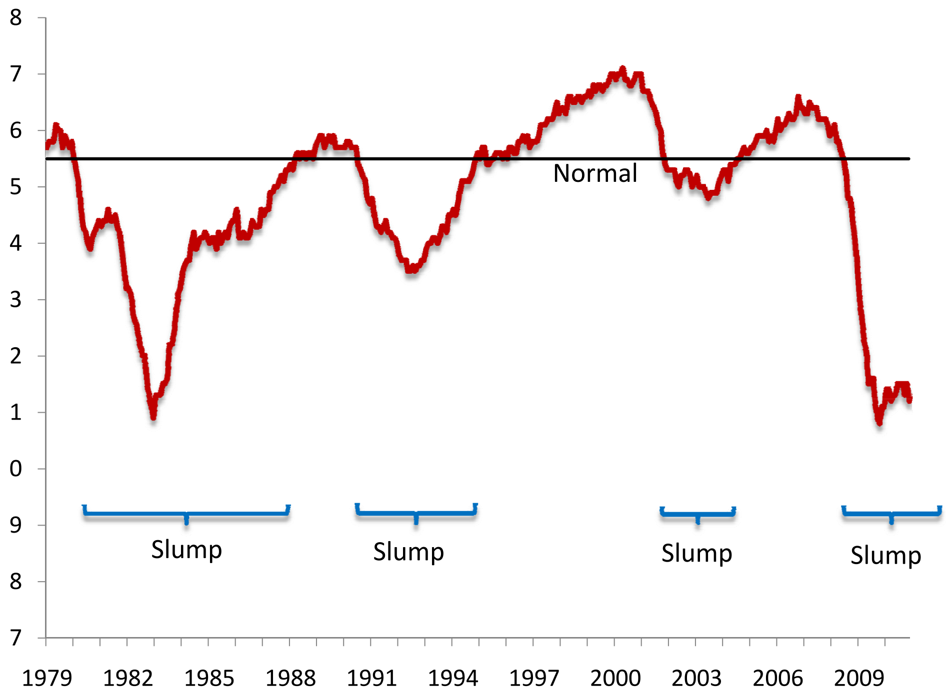

Figure 1: U.S. Unemployment Rate, 1980 through 2010

1 Introduction

At the beginning of 2011, the U.S. and many other economies are in slumps—employment is low and output is well below its growth track. The worst slump in U.S. history was the Great Depression, in which the economy contracted from 1929 to 1933 and failed to return to normal until the buildup for World War II. The slump that began with the recession at the end of 2007 will last for most of a decade, according to current forecasts. Figure 1 shows the employed fraction of the labor force aged 25 through 54 since the beginning of 1979 (the remaining fraction of the labor force is unemployed). Slumps are identified as periods when this measure of employment was less than its normal level since 1948 of 95.5 percent of the labor force.

A slump begins with a contraction, usually fairly brief, at least in comparison to the extended period of slow growth that follows the contraction. Relative to the vocabulary of peaks and troughs, a slump lasts from the time when employment falls below its normal level during the contraction to the time when employment regains its normal level during an expansion. Thus a slump spans the trough date. Usually most of the slump occurs after the trough, during the period of inadequate but positive growth. Everybody but business-cycle specialists uses the term recession to describe a slump.

All departures from full employment seem hard to explain by standard economic principles. Whenever unemployment is above normal, job-seekers line up outside any employer who is hiring. By standard principles, employers ought to be able to make mutually beneficial deals with the unemployed. As they do so, unemployment should return to normal. The process shouldn't take any longer than it takes a farmer to market an unusually large crop. The U.S. labor market demonstrated that it was capable of huge amounts of hiring without significant unemployment when it absorbed 8 million people from military service at the end of World War II.

The prolonged period of substandard growth following a major contraction seems the hardest part of a slump to explain. It's plausible—though still hard to model—that a big shock like the one that hit the U.S. economy in September 2008 will cause an immediate contraction, but harder to understand why it impeded growth five years later.

Though my commentary will generally stick to the U.S. experience, slumps have occurred in advanced countries all around the world. Japan's experience since 1990 is an extreme and interesting case.

I will adhere fairly strictly to one major restriction of the scope of this analysis. I do not discuss to any great extent the events that occurred prior to the sharp contraction of the U.S. economy at the end of 2008. My main emphasis is on the long duration of the slump, not on the dramatic developments that occurred at the outset of the slump. Many economists have told that story already.

The adverse forces that I consider are:

- The buildup of high stocks of housing and consumer durables during the 2000s
- The corresponding growth of consumer debt
- Rising financial frictions following the crisis

An important influence lies behind all three of these forces—the run-up and collapse of house prices. In effect, I am treating the residues of that influence as the driving forces of the slump.

The characteristics of the economy that are prominent in the analysis and depart from standard principles of economics are

- Real asset returns are sticky and fail to fall enough to clear the intertemporal output market
- Real wages are sufficiently sticky to result in large increases in unemployment when there is excess supply in the current output market
- Many households are financially constrained—they would be further in debt if lenders cooperated

I build a fairly simple dynamic model that embodies these characteristics. I use the model to study the response of the economy to the adverse forces that combined to create the current slump.

The starting point of the model is that, in a slump, the intertemporal market for goods now and goods later does not clear, but rather has excess supply of goods now and excess demand later because real asset returns are too high. The failure for real asset returns to fall to a low enough level results in rationing of customers to suppliers. If asset returns did fall, consumers who are not financially constrained—those who hold positive positions in financial assets—would alter their plans to consume more immediately and less in the future. The intertemporal market would clear and the excess supply of current goods that causes excess unemployment would disappear

Excessively high real asset returns result from the option open to all investors to hold cash at a zero nominal rate. That investment guarantees a riskless real return of the negative of the rate of inflation. If the risk-adjusted return in the stock market threatened to fall below this level, investors would sell the stocks and keep the cash. Because somebody has to own every share in the stock market, the result is an immediate decline in stock prices followed by enough capital gains to keep the risk-adjusted return competitive with the alternative of holding cash.

The biggest danger is serious deflation. When prices fall 10 or 20 percent per year, as during the period from 1929 to 1933, the real asset return competitive with the return on cash is far, far above the full-employment real rate and high unemployment would result even without any other adverse force.

The decline in inflation during the current slump has been about one percentage point, from around two percent through 2007 to one percent ever since. Extreme inflation stickiness in the face of large amounts of slack has saved the economy from a repetition of the experience of the Great Depression. But inflation of around one percent implies that the risk-adjusted real asset return cannot fall below minus one percent, which is well above the full-employment level when important negative forces are at work. I estimate that the normal safe short-term real rate is only 12 basis points. A decline in the equilibrium real rate by only 112 basis points below normal creates havoc, as I will show.

A substantial buildup of housing and consumer durables stocks occurred during the middle of the 2000s. I estimate that the real stock of durables (a term I will use to include housing and consumer durables) was 14 percent higher in 2007 than it would have been if the conditions of the period from 1990 to 2000 had prevailed in that year. Thus, I take the initial level of the stock of durables to be 14 percent above its stationary level when solving the model. The high initial stock causes real asset returns to fall below their stationary level. If inflation is close to zero or negative, the bound on the real returns could bind at the outset. That economy would start with excess unemployment.

Another state variable, consumer debt, rose during the 2000s. Durables secured most of the debt. Households had dissaved by taking on debt and using the proceeds to buy durables—new houses and new cars. The model starts with debt above its stationary value. As families work off their unusually high holdings of durables, they save by paying off the corresponding debt.

A third state variable, the stock of business capital, appears to have been at its stationary value in 2007. There was no impetus toward recession from business investment. Though investment fell during the recession, the model interprets this response as induced by a recession whose origins lay elsewhere and not the result of a business-related state variable that starts at a value different from its stationary value.

Starting with these state variables in 2007, the model goes into immediate recession—output is low and unemployment is high. The Fed tries to offset the recession by lowering the interest rate but cannot do more than drive its policy rate to zero, which leaves the economy still in a condition where people want to produce now but purchase later. The bulging stock of durables discourages further purchases and the overhang of consumer debt precludes higher purchases for a substantial fraction of households.

The model explains how a prolonged slump could occur. The economy remains below full employment as long as the excess durable stock and accompanying debt persists and financial frictions caused by depleted equity in financial intermediaries and households remain high.

Stripped to its basics, the story is that in the years leading up to the slump, households borrowed much more than usual and used the proceeds to buy more houses and cars than usual. They used the credit market to speed up purchases—they bought first and planned to save later. The financial crisis resulted in tighter credit standards, especially for households, and in higher spreads between borrowing rates and the cost of capital to intermediaries. In an economy without a floor on real asset returns, households not subject to borrowing constraints would have taken over purchasing unusual volumes once the constrained households reached their saving phase. But the returns could not fall far enough to induce that response. Instead, output fell and unemployment rose.

With respect to house prices, I believe that the two housing-related state variables, the housing (durables) stock and the legacy of debt-service commitments, effectively take into account the ways that the earlier housing bubble contributed to the extended slump. Households built up debt, a state variable. They also purchased more houses and other durables, which is captured in a state variable. The model has a small amount of house-price inflation that does not do justice to the actual path of house prices, which declined during 2008.

American households lost many trillions of dollars of wealth during the crisis and only a part of the loss had been recovered two years later. Many economists have argued that the slump is easy to explain conditional on the known loss of wealth: Consumption depends on wealth. When wealth falls, consumption falls. With lower product demand from consumers, the economy has fewer jobs and higher unemployment.

Explaining the inadequacies of this diagnosis of the slump is a good way to introduce the analysis that follows. The determination of consumption in a modern macro model works as follows: Households have beliefs about their future cash receipts and about current and future returns to savings. They consider consumption trajectories that balance a desire for immediate consumption against those returns. They pick the best trajectory they can afford into the indefinite future.

This process makes no direct reference to wealth. It looks past wealth to the underlying cash flows that wealth provides. Although the model implies values of wealth, the structural relation between wealth and consumption is complicated and dynamic. It is not as simple as making consumption a constant fraction of wealth. Most importantly, if, thanks to some kind of asset-price bubble, people think that they are wealthier, but do not ascribe that wealth to any future cash flow, the wealth will not affect their consumption.

Some economists have introduced a loss of wealth into accounts of contractions in a way that involves actual losses of cash flows, by hypothesizing a loss of physical business capital. An obvious defect of this approach is its unrealism—one thing that did not happen in 2008 was any known loss of plant and equipment. Starting the model in this paper with a shortfall of business capital triggers a boom, not a slump. It raises asset returns and helps the economy escape from an excess supply of current output. Spending rises as the economy regains its normal volume of capital.

There is no scope for monetary policy in a fixed-price economy. In a different economy where inflation responds to the amount of slack, a central bank stabilizes inflation by adjusting the amount of slack. Today we usually portray the policy as a Taylor rule.

The deep issues in constructing models with flexible but determinate prices given a monetary-policy rule are definitely beyond the scope of this paper. I invoke sticky money prices of output to rationalize treating the rate of inflation as exogenous. I will present evidence on recent price stability that gives me comfort in this assumption.

2 Excessive Real Returns

2.1 Nominal returns

Riskless short-term nominal interest rates dropped close to zero in the United States and many other countries in late 2008. The three-month Treasury bill rate in December 2008 was around 5 basis points and rose only to __ basis points in January 2011. The overnight federal funds rate has been equally low. Futures markets and forecasters call for a continuation of rates close to zero through 2011 and possibly 2012 as well. These returns are close to the zero nominal return on cash.

Standard principles of finance ordain that real short-term holding returns for other assets—safe longer-term debt, junk bonds, stocks, land, whatever—are the sum of the short-term safe real rate and a risk premium. These real returns are at their lower bounds when the safe nominal rate is zero. Those lower bounds are the risk premium associated with the asset class plus the negative of the rate of inflation.

2.2 Real returns

The term *real interest rate* is generic for a concept more precisely identified as an *own rate*. The latter is a rate measured in the physical units of some product. The own rates that matter in this analysis are those for consumption goods and services, for durables including houses, and for business capital. In the model I denote the own rate on output as r, which I call the real rate, and introduce the prices of other products relative to output as appropriate to generate the corresponding own return. I denote the safe nominal rate as r_n but for most of the analysis take $r_n = 0$. I call r the real rate. It is the return measured in output units available from a one-period investment at the safe nominal rate:

$$r_t = (1 + r_{n,t})\frac{p_t}{p_{t+1}} - 1, \tag{1}$$

where p is the dollar price of output.

Other real returns follow the same rule, all applying the same rate of inflation. I neglect any changes in risk, so when I assume that the safe short-term real return is roughly constant, I am assuming that returns for all assets of whatever risk are similarly constant. Although there is good evidence against the hypothesis of constant risk premiums, I do not believe that there is evidence of any significant difference between risk premiums in prolonged slumps and in normal times. In particular, slumps are not times of high volatility in asset markets or elsewhere in the economy.

2.3 Inflation

What difference does it make in a macro model if the money price of goods is fixed? Changes in the price level have important distributional effects. The rising real value of nominal debt during the Depression had important adverse consequences. But by far the most important result of fixed prices—at least from the perspective of the model in this paper—is its effect on interest rates. When the public believes that the price level is fixed now and in the future (possibly along a fixed growth path), nominal and real interest rates are locked together. The lower bound of zero in the nominal interest rate becomes a lower bound of the negative of the rate of inflation. When inflation is negative, the bound on the real rate is a positive amount. When inflation is at small positive levels, as in the past few years, the real interest rate cannot drop below a small negative level.

In the model I construct, the price *level* is immaterial. A slump is not a time when the price level is too high. Price-level irrelevance is obviously the result of drastic simplifications—ones that probably make sense only in an economy with a long history of low and stable inflation. In the model, only the rate of change, the inflation rate, matters. The key sticky price is the real interest rate. In normal times, the real rate clears the output market. If an excess supply of current goods threatens, the real rate declines. Consumers raise current consumption, especially purchases of durable goods, because they have become cheaper in relation to future goods. Businesses speed up purchases of plant and equipment for the same reason.

When the short-term safe nominal rate is zero, the corresponding real rate is

$$r_t = \frac{p_t}{p_{t+1}} - 1, \tag{2}$$

so at a constant rate of inflation π, with

$$p_t = \left(\frac{1}{1-\pi}\right)^t, \tag{3}$$

the real rate is

$$r_t = -\pi. \tag{4}$$

Thus everything turns on the rate of inflation. If inflation remains high in a slump, the real rate will be negative and thus helpful in stimulating current spending. If significant deflation breaks out, the economy is condemned to a high real rate, even higher risky rates, and severe slump.

The U.S. entered the slump with a history of low and stable inflation. Price increases were concentrated in a narrow band around two percent per year since the mid-1990s and were not much higher starting in the mid 1980s. An immediate issue following the sharp contraction in the last quarter of 2008 was whether inflation would fall and even turn into deflation as a result of the extreme slack that developed quickly. The answer, luckily, was no.

The Great Recession brought slacker product markets to the U.S. economy than had existed at any time since the depression in the 1930s. A line of thought rather deeply embedded in macroeconomics holds that product prices fall in slack markets. The logic is that sellers have much to gain by increasing output when output is low. On the reasonable assumption that marginal-cost curves slope upward, a contraction in output will cause a price-setting firm, irrespective of the amount of its market power, to cut its price in an attempt to take business away from its rivals.

Recent experience requires a fundamental reconsideration of the views that producers find it desirable to expand output by cutting prices. Their behavior across all industries suggests, to the contrary, that price-cutting is not the answer to any problem they perceive in a time of extreme slack.

Monthly inflation rates have a good deal of noise in them, from components of the price index with volatile prices, notably petroleum products. Practitioners have come up with a variety of ways of extracting a less noisy inflation signal from the monthly data. One approach is time aggregation—using annual or other multi-month changes. Another, currently the most widely used, is to study core inflation, price changes excluding the volatile food and energy components. The third—the one I favor—is to use inflation forecasts. The volatile components lose their unpredictable noise components but are not completely neglected in this approach. For the present purposes, forecasts seem the desirable approach, because it is expected future inflation that matters for the real rate.

Figure 2 shows the one-year-ahead forecast of the GDP deflator from the Survey of Professional Forecasters, maintained by the Philadelphia Federal Reserve Bank, along with the unemployment rate. The period covered starts in 1987, the year that Alan Greenspan took command of the Federal Reserve. It contains three contractions marked by rapid increases in unemployment. Inflation fell in all three, but separating the response to slack from other determinants is a challenge. The decline in inflation was greatest in the slump of 1991 through 1994, but the decline continued at about the same rate after the slump turned into a remarkable boom. Inflation did not flatten until unemployment reached 4.3 percent in 1999. Inflation was close to flat in the slump from 2001 through 2004, with just a hint of decline during the period when unemployment was rising. Finally, in the current slump, inflation took a discontinuous drop of about one percentage point early in the contraction, when unemployment was still fairly low, then stabilized at just over one percent per year when unemployment skyrocketed to the 9.5 percent level. Despite concerns that continuing extreme slack might result in further declines toward deflation, the rate of inflation has remained remarkably stable at around one percent during the recent stable period.

The concept of the non-accelerating-inflation rate of unemployment or NAIRU has had a firm grip since Friedman (1968) formulated the concept, though he called it the natural rate. The idea is that there is a critical unemployment rate such that inflation will become greater and greater if unemployment is below the rate. Today, the relevant version is the

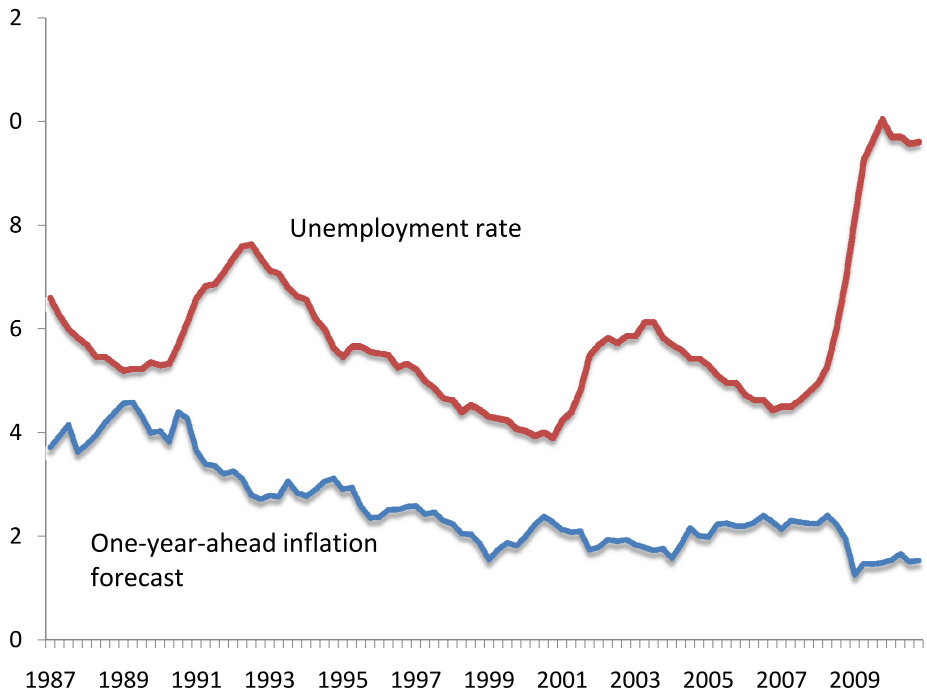

Figure 2: One-Year-ahead Inflation Forecast and Unemployment Rate, 1987 through 2010

non-decelerating-inflation rate of unemployment. By the theory underlying the concept, the rate of change of prices should fall more and more if the unemployment rate is above the critical rate. Generally the NAIRU is found to be around 6 percent. By this influential body of thought, month after month of unemployment over 9.5 percent should bring more and more deflation. Fortunately, the theory is wrong.

It is not news that NAIRU theory is a failure. Stock and Watson (2010) report that the best way to characterize the relation between inflation and unemployment is to measure downward pressure on inflation as the difference between the current unemployment rate and the lowest rate experienced in the previous 11 quarters. Thus, once a slump has lasted 11 quarters at, say, the same rate, no matter how high, unemployment loses its deflationary effect. That is exactly the opposite of the NAIRU theory. The finding is plainly consistent with the data in Figure 2. It's also plainly the case that the negative effect of unemployment on inflation is small even during the time when it has any effect.

The limited response of inflation applies at the level of components of output. Figure 3 shows annual rates of change of output and price for a number of components of GDP, over the two-year period from 2007 Q4 to 2009 Q4. The points lie along a line with a slightly positive slope—the line connecting the left-most observation to the right-most has a slope

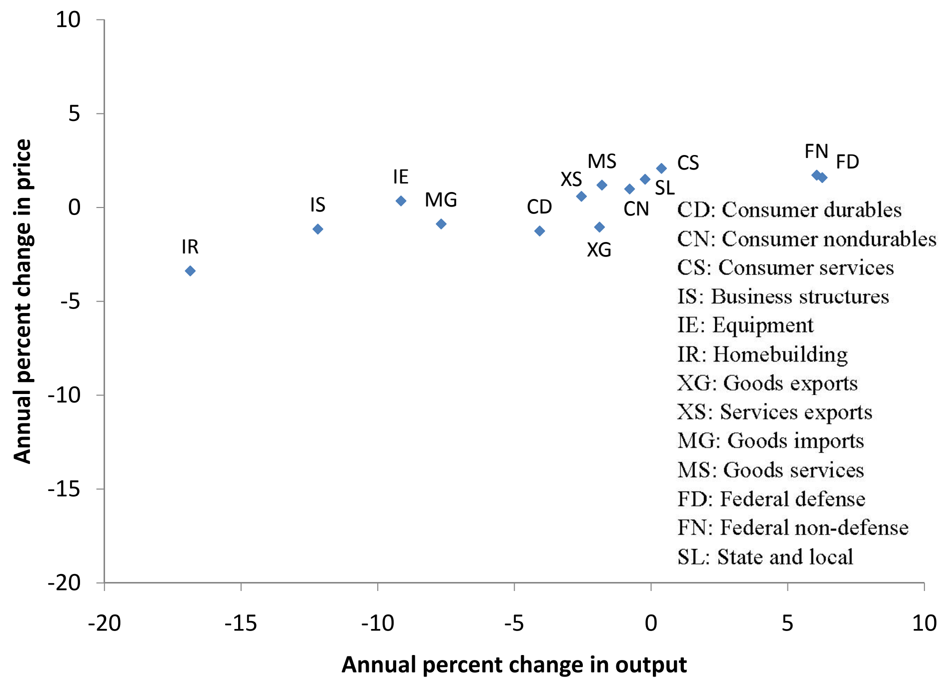

Figure 3: Annual Percent Changes in Output and Prices, 2007 Q4 to 2009 Q4

of 0.22 percentage points of price change per point of output decline. The most informative observation is for residential construction, where output declined at a 17-percent annual rate but price declined by only 3.4 percent per year. Construction is a good test case, because existing theories of sticky prices do not seem to apply to this component.

An adverse shift in the terms of trade may be an influence favoring unresponsive prices. If an increase in input prices occurs at the same time that product demand falls, product prices may hardly move at the same time that output falls. A spike in oil prices occurred in the summer of 2008. But the spike reversed by the end of 2008 and there was no meaningful shift in the terms of trade during the two years included in Figure 3. The ratio of the price indexes for imports and exports fell by 17 basis points per year during the period.

Most economic models of pricing derive a reasonably stable markup of price over cost. The dominant model of inflation embedded in practical macro models today hypothesizes that firms would like to set prices according to a markup theory, but only do so at random times. These models are inconsistent with the evidence above, because they imply that the NAIRU principle holds. They cannot explain the stabilization of inflation at positive rates in the presence of long-lasting slack.

The markup principle itself is on weak foundations. Evidence from American and Canadian car prices is highly informative about the response of retail prices to changes in cost. Thanks to NAFTA, the production of cars in North America is thoroughly integrated. The cost of building a car is the same whether sold in Canada or the U.S. Most cars have components from both countries (as well as Mexico and Taiwan) and many are assembled in both countries, with little correlation of location of production with country of sale. The Canadian dollar production cost of a car sold in Canada is very close to the exchange rate times the American dollar cost of a car sold in the U.S. Standard theories of stable markup ratios for oligopoly products suggest that an appreciation of the Canadian dollar relative to the American dollar will cause a combination of an increase in the American dollar prices of cars in the U.S. and a decrease in the Canadian dollar price of cars in Canada so that the markups remain constant. Thus the Canadian price of cars translated into American dollars at the exchange rate should be stable in relation to the American price. Figure 4 tests this hypothesis. Far from being stable, the ratio of the U.S. dollar translation of the Canadian retail price to the American retail price, shown in the heavy red line, is highly volatile and persistent. Further, the movements of the price ratio are highly correlated with the movements of the exchange rate. When the Canadian dollar is strong, the profit margin for sales of North American cars in Canada widens in proportion. When it is weak, car makers continue to sell in Canada at very low margins. The phenomenon of stabilization of prices against cost changes associated with exchange rates is well known (but not well explained) in a large literature in international trade.

To formalize the analysis, let p_C be the Canadian price of a car in Canadian dollars and p_A be the American price of a car in American dollars. Further, let c be the marginal cost in euros common to cars sold in both countries and let $x_{C,E}$ be the Canadian dollar value of a euro and $x_{A,E}$ be the American dollar value of a euro. Define the markup ratios of price over marginal cost as m_C and m_A. Then

$$p_C = m_C x_{C,E} c \tag{5}$$

and

$$p_A = m_A x_{A,E} c. \tag{6}$$

The ratio of the prices is

$$\frac{p_C}{p_A} = \frac{m_C}{m_A} x_{C,A}, \tag{7}$$

where $x_{C,A}$ is the Canadian dollar value of an American dollar. By standard principles, the markup ratios should be constant and so their ratio should be constant. In that case, the

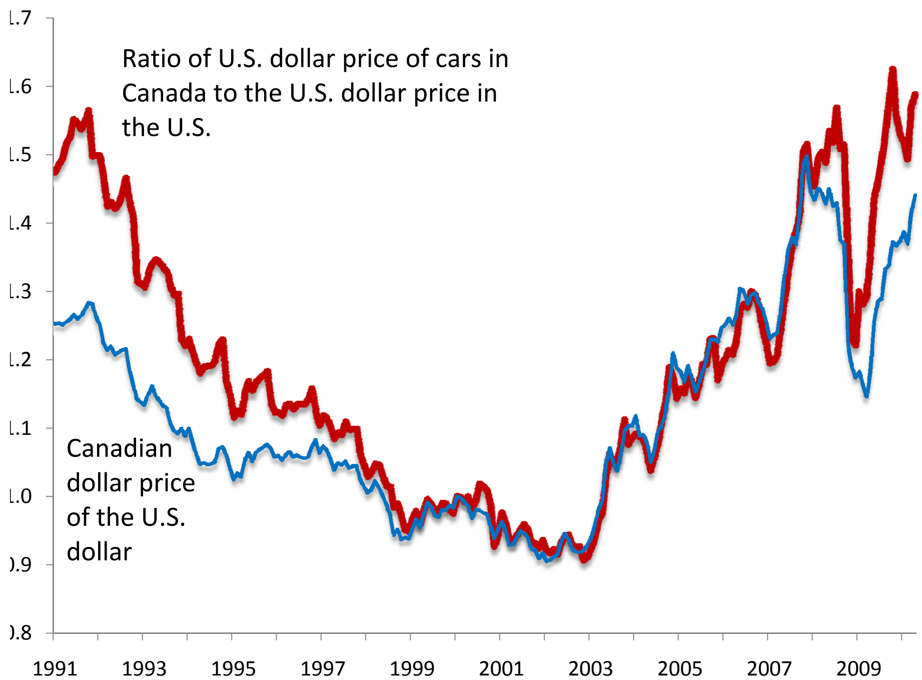

Figure 4: Canadian and American Car Prices and the Exchange Rate, 1991 to 2010

price ratio should track the exchange rate $x_{A,C}$. On the other hand, if car prices are sticky in each country's currency, $\frac{p_C}{p_A}$ will be close to constant.

A specification that nests the two cases is

$$\frac{p_C}{p_A} = \mu x_{C,A}^{\pi}. \tag{8}$$

If the markup principle holds, the prices will move in proportion to the exchange rate and μ will be the constant ratio of the markups and π will be one. If prices are completely sticky in the respective currencies, π will be zero and μ will be the constant ratio of the prices. On the reasonable hypothesis that the exchange rate is uncorrelated with the disturbances in car prices in the two countries, I estimate π by regression in logs with correction for first-order autoregressive errors. The estimated value is 0.045 with a standard error of 0.053. The hypothesis of entirely sticky prices is easily accepted while the hypothesis of a constant markup ratio is overwhelmingly rejected.

2.4 Wage inflation

Nominal wages were even less responsive to the development of huge slack in 2008 and 2009. The best measure of hourly wages, inclusive of benefits and all other types of compensation, is

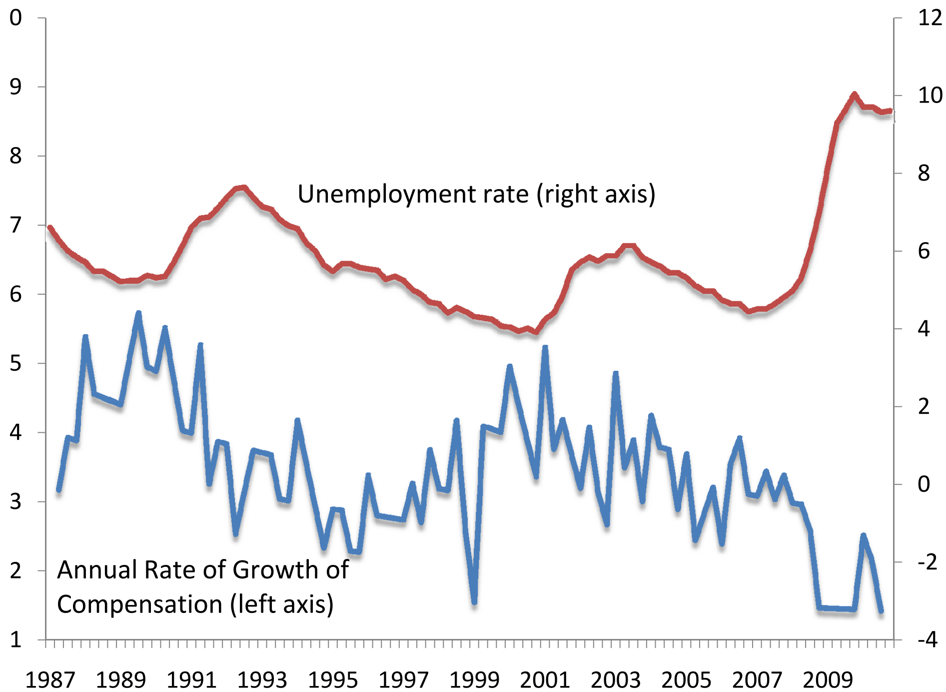

Figure 5: Wage Inflation and Unemployment

the Employment Cost Index of the Bureau of Labor Statistics. Figure 5 shows the quarterly changes in nominal compensation starting in 1987, along with the unemployment rate. Wage inflation fell during the three slumps in that period. In the slump of 1991 through 1994, the drop occurred at the beginning of the slump and wage inflation stabilized during the rest of the slump. In the slump of 2001 through 2004, wage inflation declined along a smooth path, but the rate of decline was the same during the slump and during the boom that followed it. In the current slump, wage inflation dropped by about a percentage point at the beginning and has been stable ever since, except for one quarter in which it jumped up by a percentage point.

Wage inflation appears to behave in much the same way as price inflation. During the sharp contraction at the beginning of a slump, wage inflation falls a bit. During the period of extended high unemployment for the rest of the slump, wage inflation remains constant. The behavior of wage inflation plainly refutes the NAIRU hypothesis.

As with price inflation, wage inflation is unresponsive to slack in the cross section of industries. Figure 6 shows the annual changes in employment and hourly wages for an 8-industry breakdown of private employment. The line is almost exactly flat. The slope is 3 basis points of wage change per percentage point of output change. Again, construction

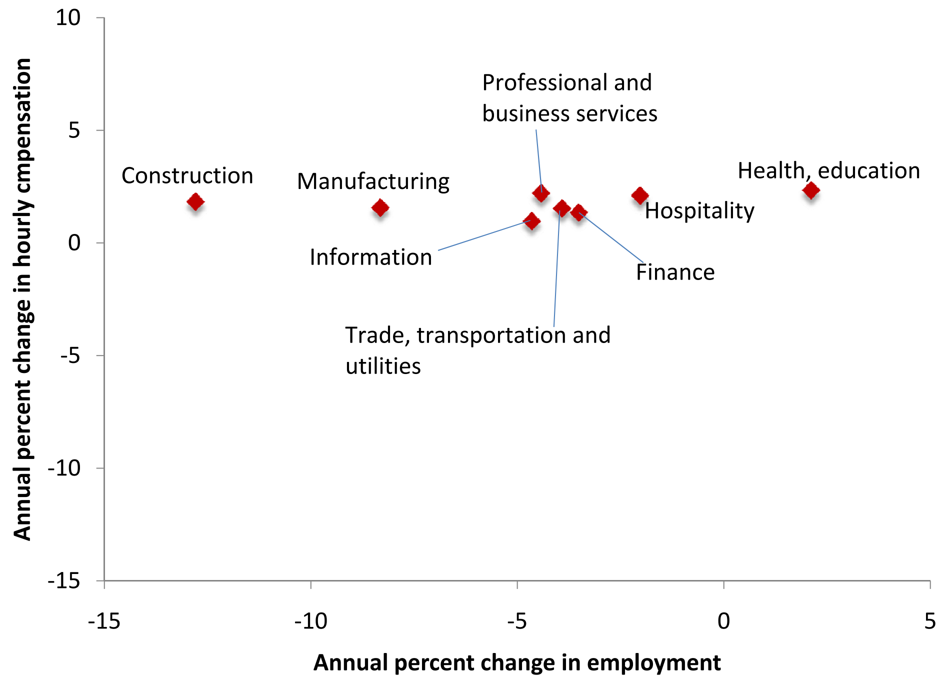

Figure 6: Annual Percent Changes in Employment and Hourly Compensation, end of 2007 to end of 2009

is the most informative data point. Compensation per hour *rose* at the same time that employment was falling by 13 percent per year.

Macroeconomists need to redouble their efforts to understand the cyclical behavior of prices and wages, especially the lack of decline of inflation during all but the beginning of a slump.

2.5 Implications for real returns

Apart from a small decline at the beginning of a slump, inflation appears to remain the same for the duration of the slump. For as long as the short-term safe nominal interest rate remains pinned at zero, the risk-free short-term real rate remains fixed at minus the rate of inflation. Under the hypothesis that risk premiums remain approximately constant during the quiescent period of a slump following the disturbances that cause the initial contraction, the finding of fixed inflation implies that all real asset returns are roughly constant. They are higher than they would be absent the option for investors to hold cash.

Thus the macroeconomics of slumps focuses on the consequences of real asset returns that are too high to clear the market for current output.

3 Rationing of Customers to Sellers in the Output Market and Resulting Unemployment

Unresponsive prices and rationing of sales opportunities in the output market are closely connected topics. Economists believe instinctively that product prices should fall when businesses are capable of producing more output than they are currently selling. But in the modern environment, businesses do not seem to believe that undercutting their rivals by significant amounts is the right way to deal with low sales volumes. A fragmentary literature in macroeconomics and, almost independently, in marketing science, explains why price cuts may not be in the interest of sellers in particular settings. None of this research has come close to explaining the behavior of the aggregate price level, which includes many products with volatile prices set in open markets. Yet the aggregate rate of inflation is unresponsive to slack except early in contractions, and even then, the response is remarkably weak.

When an economy is in a slump, firms could seemingly profit by taking sales away from rivals. In the model, I adopt a specification that implies that price cuts for this purpose are small and temporary. At this point, I lack a convincing foundation for the specification. As discussed above, price stickiness takes the form of a predetermined price trajectory that grows at a low, constant rate of inflation. I assume that firms adhere to sticky prices on this trajectory as long as customers are not rationed, but deviate below the sticky price path when rationing occurs, in efforts to attract more customers from other sellers. The yield of new customers from a price deviation is lower when the rationing is tighter—that is, when the gaps between sticky real asset returns and the equilibrium real returns are greater. The marginal revenue product of labor is correspondingly lower. Thus when the excess supply of goods is high, the benefit of hiring another worker is low, because it is hard to find a customer for the worker's output. In effect, market power rises in that situation. Cyclical variations in the markup of price over cost are a feature of some New Keynesian models—see Hall (2009a)—but the rationalization is somewhat different here.

I adopt a really simple version of the Diamond-Mortensen-Pissarides (DMP) model to describe the labor market. In the standard version of that model, as in Mortensen and Pissarides (1994), the payoff to hiring an additional worker is that worker's marginal revenue product. The source of variation in the marginal revenue product generally cited in the DMP literature is productivity. As a result, the model is vulnerable to the critique that declines in productivity do not seem to be the driving force of recessions, at least since the 1980s. In the model of this paper, the marginal revenue product falls when there is rationing of customers. The way that the DMP model may be able to generate fluctuations in unemployment driven by changes in the benefit to the firm from hiring added workers is well understood. Shimer (2010a) deals extensively with the conditions under which the model delivers an unemployment rate immune from aggregate influences. If the bargained wage is less flexible than under those conditions, cyclical unemployment fluctuations will occur.

As Shimer (2005) demonstrated, the Nash wage bargain in the standard DMP model precludes significant fluctuations in unemployment. When the marginal revenue product of labor falls, the wage falls virtually in proportion, so the incentive to hire remains essentially the same and unemployment remains at its natural rate. I adopt a simple framework that implies less wage flexibility, along the lines of Hall (2009b). Thus the framework includes both sticky prices and sticky wages.

4 Implications of Sticky Real Asset Returns

I have formulated the implications of sticky inflation and the investment option of cash as resulting in sticky and excessive real asset returns. Most of the literature on this topic adopts the equivalent formulation of a zero lower bound (ZLB) on the nominal interest rate.

The basic intuition describing the operation of the ZLB economy is straightforward—see Krugman (1998) in the context of Japan in the 1990s and Christiano, Eichenbaum and Rebelo (2009) in the recent U.S. context. The bound keeps the real interest rate above its equilibrium level. Households respond by offering to work harder in the current period but to defer consumption to the future. Investment is low because of the high real rate. The economy suffers from an incipient shortfall of investment below saving. Income has to fall enough to bring saving into equality with investment. Some adverse shock has brought the economy into the ZLB state but the bound amplifies its effect. Or, to put the point in reverse, a stimulus, such as an increase in government purchases, has a bigger effect on output in the ZLB economy than in a hypothetical alternative economy where the nominal interest rate can be negative.

Recent work on ZLB macroeconomics has been, without exception as far as I know, embedded in the standard New Keynesian framework whose canon is Christiano, Eichenbaum and Evans (2005). That model is complicated because of its objective of matching many of the dynamic features of the economy. Christiano, Ilut, Motto and Rostagno (2007) and Christiano, Trabandt and Walentin (2010) add an important feature missing from earlier New Keynesian models, an explicit treatment of unemployment.

4.1 The excessive return on cash

The source of sticky and excessive real asset returns or the ZLB is the government's willingness to issue a security—currency—that pays a needlessly high return. With the nominal interest rate at zero, no private bank would issue currency, because the real return to issuing currency—the nominal interest rate less the cost of printing and managing the currency—is negative and thus commercially impractical. Absent the government's willingness to issue currency, an economy could function with a negative nominal interest rate. Currency would disappear or would be subject to holding charges. The government would charge banks for the privilege of holding reserves whose real values would rise over time. Buiter (2009) discusses these issues in detail, with cites to recent and vintage writings.

Does the government really need to offer unlimited quantities of currency to investors? One alternative would be to limit the quantity of currency by reducing the monetary base. That step would be a standard monetary contraction. It would raise the nominal interest rate to positive levels and worsen the slump. The second alternative would be to discontinue the standard policy of every central bank to trade reserves for currency at par in unlimited amounts. Currency would then trade at a premium. Banks would charge \$110 to a depositor's account for issuing \$100 at the ATM. Neither of these alternatives seems likely to gain acceptance. In a slump, the central bank will continue to provide currency despite its harmful effects on real activity and its fiscal burden.

The essence of the drag on the economy from the availability of government currency is that the government forces all real returns above their equilibrium levels, creating excess supply and rationing in the output market, as the excessive real returns induce workers and firms to defer demand for consumption and investment. Removal of the option to invest in government currency would call for fundamental reorganization of the monetary institutions of the U.S. and other advanced countries. That step is politically impractical, so I take the government's willingness to offer currency in unlimited quantities to those who choose to hold it as a permanent property of the economy.

To simplify the discussion and focus on the central issue, I treat money only as a store of value and not as a medium of exchange. In the model, no money is actually held, either in normal times (when its return falls short of the market real return) or when the ZLB binds. Its only role is to create customer rationing in the latter case because each household correctly perceives an opportunity to lend to the government at an above-market rate.

5 Liquidity Constraints and Loan-Service Commitments

During the 2000s, prior to the crisis in 2008, U.S. consumers took on additional debt. Mian and Sufi (2010) demonstrate large differences among states in the U.S. in durables purchases negatively correlated with indebtedness. During the slump, through the end of 2010, debt as a fraction of income declined back to its level in 2000. Tighter standards for consumer credit result in a substantial decline in household debt. The resulting decline in consumption and increase in saving was one of the important adverse forces leading to the slump.

A significant fraction of American consumers appear to be liquidity-constrained. I take a family as liquidity-constrained if its holdings of net liquid assets are than two months of income. Net liquid assets are the difference between holdings in savings accounts and the like and borrowing from credit cards and other unsecured forms. In the 2007 Survey of Consumer Finances, households illiquid by this standard earned 58 percent of all income. The fraction of households that were constrained—74 percent—is much higher because lower-income households are more likely to be constrained.

I incorporate these facts into the model by dividing consumption into two parts. Consumption of unconstrained households obeys the standard life-cycle model, while consumption of constrained households is their earnings less their payments on outstanding credit, which forces them to save. To determine the amount of the required repayments, I use the following logic: Let D_t be the outstanding debt of constrained households in quarter t. Constrained households always borrow the maximum allowed, so their debt is controlled by the borrowing limits imposed by lenders.The borrowing interest rate is r_D. The sources of funds for constrained households are income $\bar{y}_t$ and increased borrowing, $D_t - D_{t-1}$. Uses of funds are consumption $\bar{c}_t$ and payment of interest on earlier debt, $r_{D,t-1}D_{t-1}$. Thus consumption is

$$\bar{c}_t = \bar{y}_t + D_t - (1 + r_{D,t-1})D_{t-1}. \tag{9}$$

I let

$$s_t = r_{D,t-1}D_{t-1} - \Delta D_t, \tag{10}$$

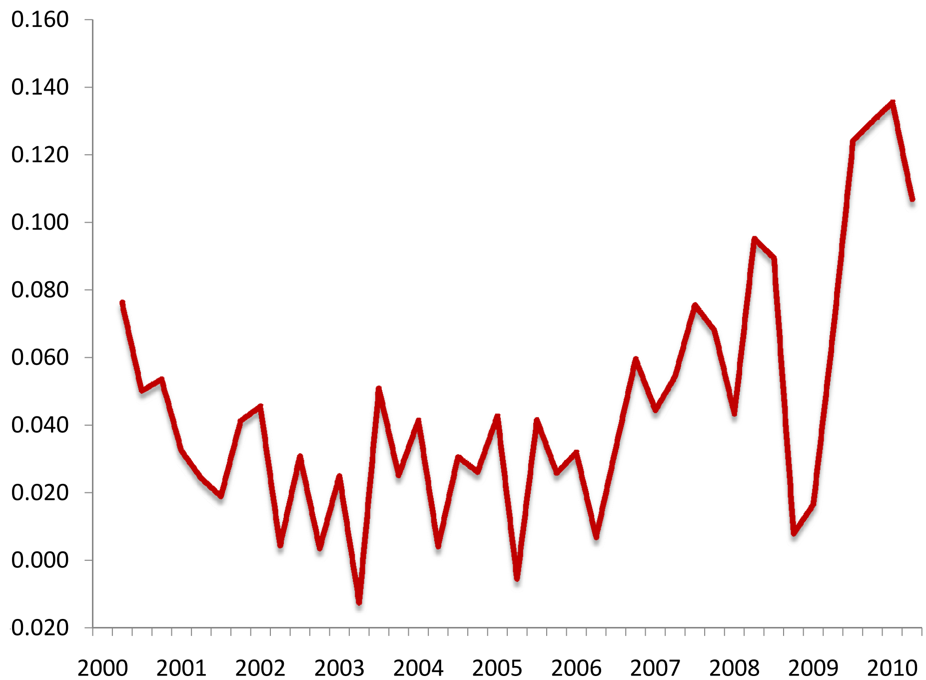

Figure 7: Burden of Debt Service, as a Fraction of GDP

the net burden on consumption relative to income associated with debt service. It is almost always a burden, because the level of debt is high enough that interest payments usually exceed new borrowing. In a stationary setting, $D_t = \bar{D}$, a constant, debt repayments are zero, and household purchases of consumption goods are less than income by the amount of interest they pay, $r_D \bar{D}$.

Figure 7 shows the burden s_t, as a fraction of GDP, calculated from Flow of Funds data on consumer debt and NIPA data on household interest payments. Prior to 2007, the burden was close to zero—new borrowing came close to covering interest payments. Tightening of credit began even before the onset of recession at the end of 2007 and continued to the end of 2010.

In a full-employment economy containing unconstrained and constrained households, the tightening of credit as shown in Figure 7 would not have contributed to a slump. Rather, the interest rate would clear the output market, thanks to the absence of any lower bound on the interest rate. When constrained households cut back consumption spending, including purchases of new houses and consumer durables, low rates would induce unconstrained households to consume more by borrowing, thus offsetting the saving of constrained households.

In the economy up to the end of 2007, consumers added substantially to their stocks of durables, mainly in new houses, but also in cars and other consumer durables. They took on additional debt in the process and became vulnerable to a tightening of credit, or, for that matter, even a stabilization of the credit/income ratio. Thus two critical state variables were at high values at the end of 2007, the durables stock and debt service commitments of constrained households.

The model starts at the end of 2007 with those state variables at the high values. I do not model the economy during the run-up, but others are working on that topic. Macroeconomists have built asset-price crashes into general-equilibrium models—Burnside, Eichenbaum and Rebelo (2010) is a recent example where beliefs about appreciation spread among homeowners like an epidemic disease. No view about the origins of bubbles or crashes is yet firmly established.

An economy starting with a stock of durables above its steady-state level and with consumers who have large commitments to save by repaying debt will reach equilibrium with a low interest rate if the rate is not bounded, or it will have excess unemployment if the rate hits the bound.

Note that the rate that matters here is the return to the saving of unconstrained households who would respond to low rates by dissaving to offset the saving of the constrained households. Consumer borrowing rates determine the burden of debt service but do not clear the output market, because constrained consumers are not marginal participants in that market.

Figure 8 shows the ratios of business capital and durables-housing to GDP since 1990. Nothing special happened to capital, but the figure shows a conspicuous bulge in durables-housing from 2000 to 2006. The ratio is 14 percent higher in 2007 than it was in 2000 and the same 14 percent higher than the average from 1990 through 2000. Accordingly, I will start the model with its durables-stock state variable 14 percent above it stationary level.

6 Financial Frictions

Hall (2010) describes a framework for studying financial frictions and Hall (2011) gives many details and cites to recent work on this topic. The framework in those papers shares many elements with the model developed here.

A widely studied setup generates frictions from the agency relationship between investors

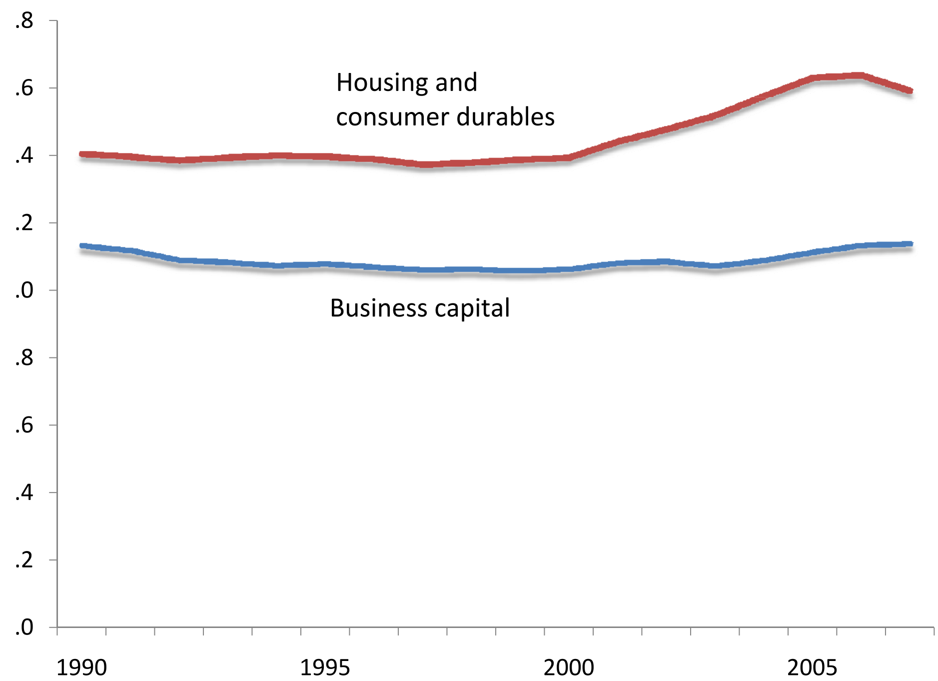

Figure 8: Ratios of capital and durables to GDP

and financial intermediaries—see Gertler and Karadi (2009) for an application to the slump. Investors lack the expertise to invest directly in productive enterprises. They place their funds with intermediaries who will abscond with some fraction of the funds unless their continuation values exceed the value of absconding. In equilibrium, the intermediary charges entrepreneurs more for credit than the amount paid to investors. The present value of the difference is the needed continuation value. The present value arises from spread between the intermediary's lending and borrowing rates and from the intermediary's capital. The setup has been influential in the theory of bank regulation, which holds that the value of a bank's franchise protects depositors and deposit insurers against excessive risk-taking by banks.

When assets held by intermediaries lose value—as from a decline in real-estate prices—the continuation value threatens to drop below the value from absconding. To prevent absconding, the investors accept a widening of the spread. Thus the behavior of credit spreads is central to this theory of variations in financial frictions.

Following this logic, I introduce a variable interpreted as a wedge between returns earned by savers and the cost of capital to businesses and households. The wedge is equivalent to a property tax on business capital and household durables, including houses. Increases in the wedge are potent sources of lower output and higher unemployment. In the presence of real asset returns made sticky by the bound resulting from the option to invest in cash (the ZLB), increases in the financial friction wedge are easily capable of creating a severe slump by themselves.

6.1 Financial frictions and interest spreads

The financial wedge connects the decline in real-estate prices that occurred prior to the beginning of the slump to the severity of the slump. Many financial intermediaries suffered severe depletion of their equity from long positions in real-estate related assets. Measuring the resulting widening of spreads is a challenge. While it is easy to measure spreads for traded instruments, those spreads are not infected by the wedge suggested by the agency theory. The spread between BAA corporate bonds and Treasurys of the same maturity widened stunningly in late 2008 but the difference did not reflect profit accruing to any intermediary. Spreads of that type came back to normal way too fast to account for any of the persistence of the slump.

One important challenge is to distinguish the part of the spread that arises from the probability of default from the part that is a true friction. The ideal measure of the spread for bank loans would be the lending rate for new loans, less the best forecast of the default probability and less the bank's cost of funds. Figure 9 shows two measures of the spread based on data from the Federal Reserve Board for business loans of $1 million or more. The total spread is the loan rate reported by banks (not necessarily limited to new loans) less the federal funds rate as a measure of the cost of funds. The total spread rose by almost two percentage points and remained persistently high. The spread net of chargeoffs deducts banks' loan chargeoffs stated as fractions of loan values. As in the 2001 recession, the spread measured this way plunged during the contraction leading into the current slump. But chargeoffs are losses on loans made in earlier years and are not a satisfactory measure of expected defaults on new loans. Contractions cause large unexpected levels of chargeoffs. I conclude that the evidence mildly supports the view that spreads of the type associated with frictions rose and remained high during the slump. I note that the total spread remains persistently high at the same time that spreads for traded securities have returned to normal.

Figure 10 shows the spreads between the rate reported by the Federal Reserve Board for credit-card borrowing and the federal funds rate, without and with the deduction for

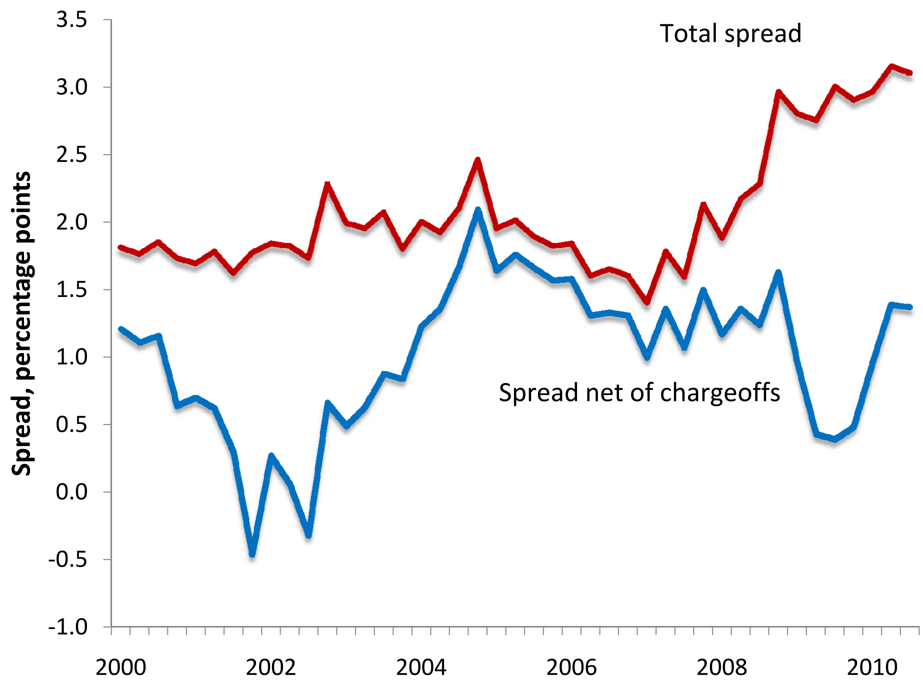

Figure 9: Spread, in Percentage Points, between Business Loan Rates and Banks' Cost of Funds

chargeoffs. The total spread is high—always above 8 percent—but it is clear that most of the spread at all times arises from high chargeoffs for credit cards, which provide unsecured credit to households at all but the lowest income levels. Chargeoffs probably lag much less for credit cards than for business bank loans, so the current chargeoff rate is probably a better measure of expected defaults for credit cards than for business loans. If so, the data for credit cards do not suggest much increase in agency frictions in consumer credit markets.

The most important component of household credit is residential mortgages. Figure 11 shows the spread between the rate reported by the Federal Reserve Board for conventional (30-year fixed rate) mortgages and the yield on 10-year Treasury notes, a reasonable match to the actual duration of mortgage debt. The total spread rose early in the slump until the Federal Reserve intervened and restored the total spread to roughly normal levels. As with business loans, chargeoffs arise from loans made in earlier years, so negative spreads net of chargeoffs are not indicative of expected losses on new loans. Mortgage underwriting practices have changed dramatically during the slump to try to limit losses on new mortgages.

My discussion of intermediation spreads has focused on the agency model, where spreads measure distortions but not actual losses of resources. Agency frictions have much the same

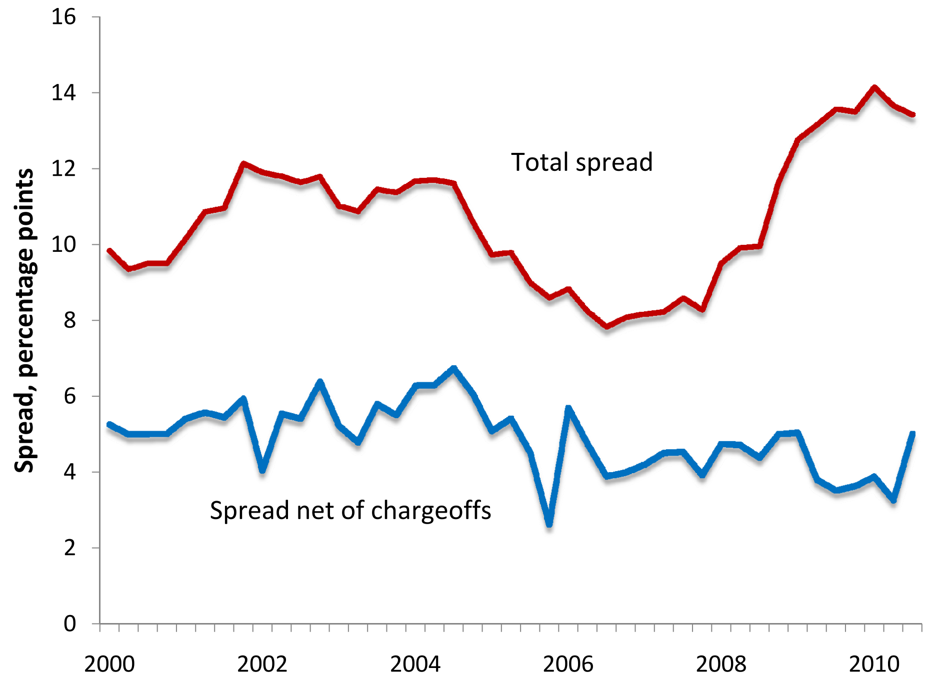

Figure 10: Spread, in Percentage Points, between Credit-Card Rates and Banks' Cost of Funds

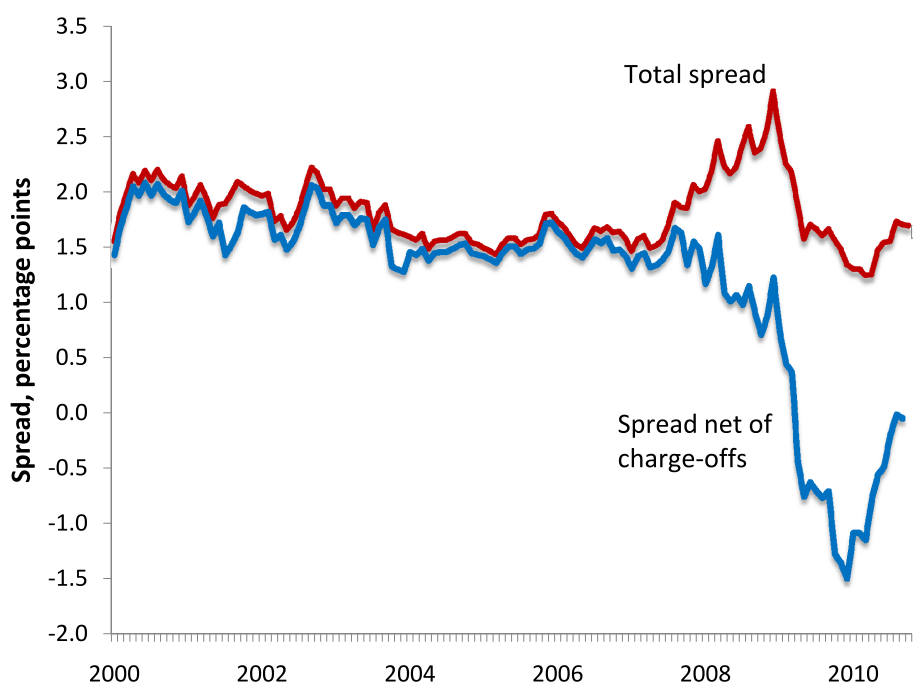

Figure 11: Spread, in Percentage Points, between Mortgage Rates and 10-year Treasurys

effect as taxes—they create wedges and resulting inefficiencies, but do not consume output. Chargeoffs are relevant apart from measuring agency frictions. In particular, resources are consumed and a different type of friction occurs when a borrower is unable to perform on a debt obligation. Another branch of the literature on financial frictions, starting from Townsend (1979), observes that debt contracts have the property that the lender need not consume resources monitoring the borrower unless the borrower fails to make good on the simple promise to repay the loan, an act that is costless to verify. When the borrower is unable to repay, the lender incurs substantial costs to recover value through a workout or bankruptcy. When borrowers suffer losses in asset values, the probability of default rises and total spreads rise. The friction is less than the spread, because one consequence of a default is a simple transfer of value from the lender to the borrower. The friction is the amount of the loss accruing to the borrower and lender jointly—it comprises bankruptcy costs, business interruption costs, and the like. But the friction is likely to move in proportion to observed spreads, so a widening of the total spread will usually indicate an increase in the financial friction.

6.2 Credit rationing

Lenders always ration credit. They need to overcome substantial adverse selection problems. They set standards for borrower eligibility and spend resources verifying that borrowers meet the standards. One of the consequences of a contraction, especially one with an initial financial crisis, is a tightening of lending standards for both businesses and households. Adverse selection becomes a more serious danger when more borrowers are close to the margin of failure.

Lending standards are increasingly based on credit scores and other metrics, but I am not aware of any systematic compilation of quantitative standards into an overall index. The Federal Reserve Board carries out a quarterly survey of senior loan officers of banks with respect to lending standards for a variety of types of loans. An example of a question in the survey is "Over the past three months, how have your bank's credit standards for approving applications for credit cards from individuals or households changed?" The permissible answers are (1) Tightened considerably, (2) Tightened somewhat (3) Remained basically unchanged (4) Eased somewhat, and (5) Eased considerably. Although the answers are qualitative, it appears possible to create an index of standards from the answers.

To this end, let x_t be an index of lending standards, interpreted as the mean across banks, where the change in the bank's own index is normal with mean $\Delta x_t + \mu$ and unit standard deviation. If a bank's own index change is in the interval $[-\nu, \nu]$, it reports that its standards "remained basically unchanged." The Federal Reserve reports the difference in the fraction of banks that reported a tightening of standards and the fraction that reported a loosening—this is called the net change. It is

$$\text{Net change } = \Phi(\Delta x_t + \mu - \nu) - \Phi(-\Delta x_t - \mu - \nu), \tag{11}$$

where Φ is the standard cumulative normal distribution (see the Appendix for details). To estimate the parameter ν, I observe that when $\Delta x_t + \mu = 0$, that is, when standards are not changing, the probability of the middle answer is the probability for $[-\nu, \nu]$, which is $2\Phi(\nu) - 1$. For October 2010, a time of small net change, the Federal Reserve reports these probabilities, which are around 0.85. The corresponding value of ν is around 1.4. I calculate the time series for $\Delta x_t + \mu$ by solving equation (11) and estimate μ as the mean of the series. The identifying assumption is that Δx_t has mean zero. My assumption that the cross-sectional standard deviation across banks is 1.0 amounts to a normalization of the units of the index.

Figure 12 shows the resulting indexes, the cumulations of the estimated changes. Because the indexes are constructed to start and end at zero, nothing should be read into the lack of trend, though it remains a reasonable assumption. Further, because each of the three indexes starts arbitrarily at zero, but in different years, the relative values of the indexes have no significance. Standards for business loans track the business cycle almost perfectly. Each recession saw tightening up to a peak that occurs around or a little after the trough of the cycle. Then standards began to ease, gradually during the 1990s and more rapidly in the 2000s. For credit cards, special factors not relevant for current purposes caused a tightening during the expansion of the 1990s. Starting from the peak in standards in 2003, credit card standards behaved similarly to business lending standards. The Federal Reserve added a question about mortgage lending standards in 2007, around the time of the low point in the other two indexes. The index of mortgage standards rose dramatically and has declined only slightly below its peak value.

Figure 13 provides further confirmation of a substantial and persistent increase in household rationing from the financial crisis. It shows an index from Google Insights of queries for the term "withdrawal penalty." The index measures, in arbitrary units, the share of all

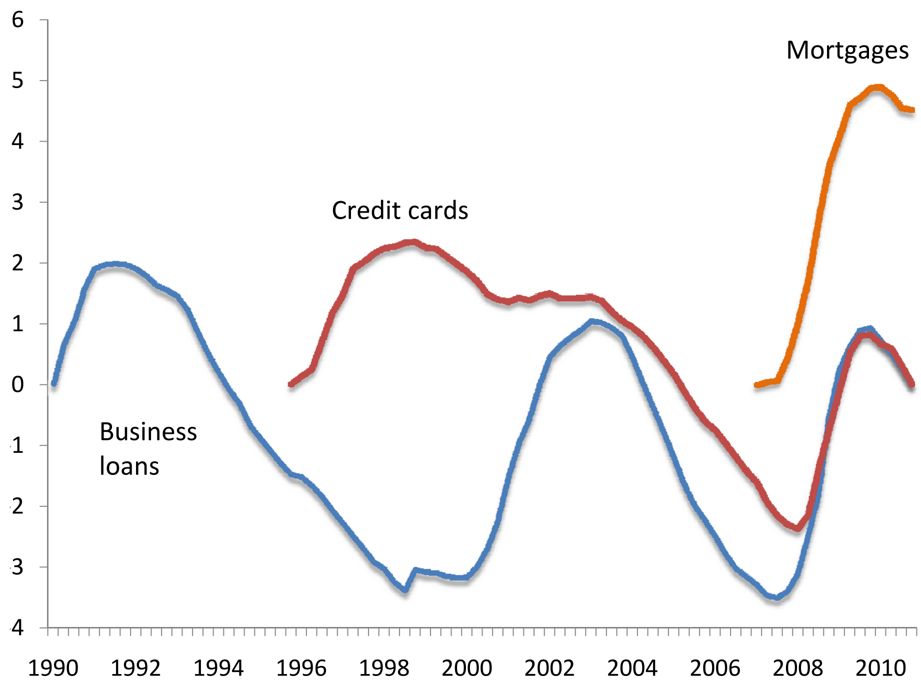

Figure 12: Indexes of Lending Standards Inferred from the FRB Senior Loan Officer Survey

search queries on Google for this term. The index jumped up in late 2008 and has remained high ever since.

6.3 Implications

Rationing through the application of lending standards enters the model in the same way as agency frictions. Borrowers behave as if credit were more costly than the interest rate they pay for the amounts they are actually allowed to borrow. The data in Figure 9 through Figure 13 suggest that agency frictions, default costs, and rationing all worsened during the crisis and have remained at levels close to their peaks during the crisis. I combine all of these factors into the wedge discussed earlier. Although the evidence seems strong that the wedge increased and has remained high, I have not found a basis for quantifying its rate. Rather, I will demonstrate that rates in a reasonable range have powerful negative effects on output and employment.

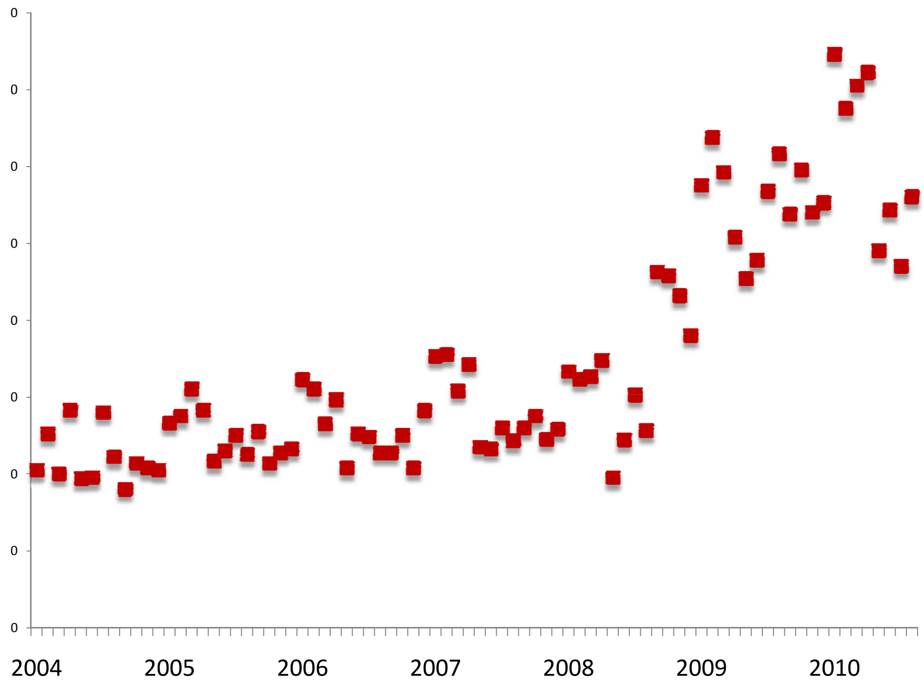

Figure 13: Share of Google Search Queries for the Term "withdrawal penalty"

7 Basic Properties of the Model

In this section, I will describe the model in terms of diagrams relating to events in the economy in the first calendar quarter of its dynamic equilibrium. The diagrams are more than schematic—they portray the actual equilibrium of the full dynamic model that I will lay out shortly.

Figure 14 shows the model's excess supply function for current output. The thought experiment behind the graph is to give the model a real interest rate—not subject to any bound—for the first quarter and ask how much output the economy will produce in that quarter relative to the amount that its participants will purchase. One can imagine that the model economy is isolated from the rest of the world in all but the first quarter, when it can borrow from or lend to other countries at a prescribed interest rate in an amount it chooses. In the following quarter, the loan is repaid but no new lending transaction is made. The graph shows that, if the rate is low, the economy chooses to borrow so it can produce less than it purchases, and chooses to lend to accommodate production above purchases if the rate is high. At the interest rate r^*, the economy has zero excess supply of current output—it produces the amount it purchases. That point is the equilibrium of the economy, which is

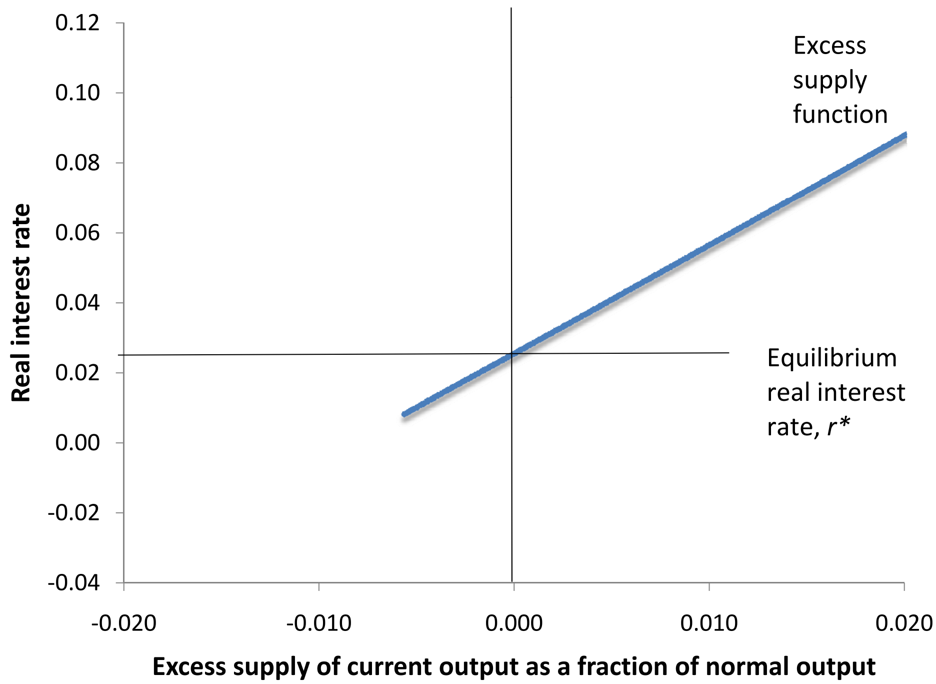

Figure 14: The Model's Excess Supply Function for Current Output

actually closed at all times.

Figure 15 introduces a lower bound on the interest rate in the same closed economy. The bound binds on the interest rate, raising it above the equilibrium level. The result is an excess supply of output. The higher interest rate induces consumers to defer consumption, especially purchases of houses and other durable goods. It also diminishes current business investment.

I do not dwell on the effects of an elevation of the interest rate when the bound suddenly materializes. As I discussed earlier, the bound is minus the rate of inflation and no important decline in inflation occurred in the slump. Rather, what matters is a shift in the determinants of the equilibrium interest rate in an episode when the interest rate is already at the bound. What happens in an economy with a fixed interest rate if the excess supply of current output changes? That is the question of the multiplier, with the interest rate held constant. Figure 16 shows the values of the reduction in output that accompanies different levels of excess supply. The slope is just over two in absolute value. Many elements of the model enter the determination of the slope. About half of consumption obeys the life-cycle principle, under which households solve a long-term optimization problem and thus respond relatively little to a transitory loss of income. The other half of consumption occurs in constrained households

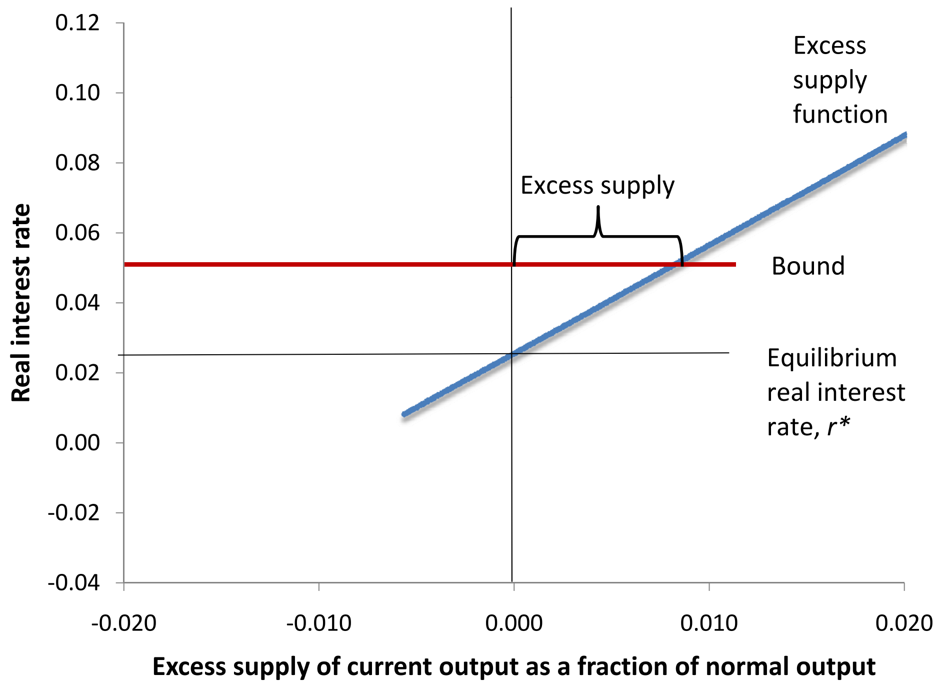

Figure 15: Binding Constraint Elevating the Interest Rate and Causing Excess Supply of Current Output

where consumption does track current income, so the standard consumption multiplier is part of the story. Investment behavior in the face of adjustment costs is another.

8 Long-Horizon Fully Specified Dynamic Model

The economy in the model lasts for many years and households last as long as the economy. Households consume nondurable goods and services and the services of durables, including housing. Output is divided among three uses: nondurables consumption, investment in new durables and housing, and investment in business capital. The length of a period is a calendar quarter. Because the real interest rate is bounded by the negative of the rate of inflation, rationing of customers may occur in the product market in any quarter. Both types of investment incur standard quadratic adjustment costs, so Tobin's model of investment applies. Household preferences have constant intertemporal elasticity of substitution, not necessarily equal to one. Firms use resources to attract customers. The benefit of hiring a worker falls when customers become harder to find, so recruiting incentives fall and unemployment rises.

Uncertainty is not an important element in the model. In particular, the model lacks any second-moment effects. Decision-makers have perfect foresight. I don't regard this as a

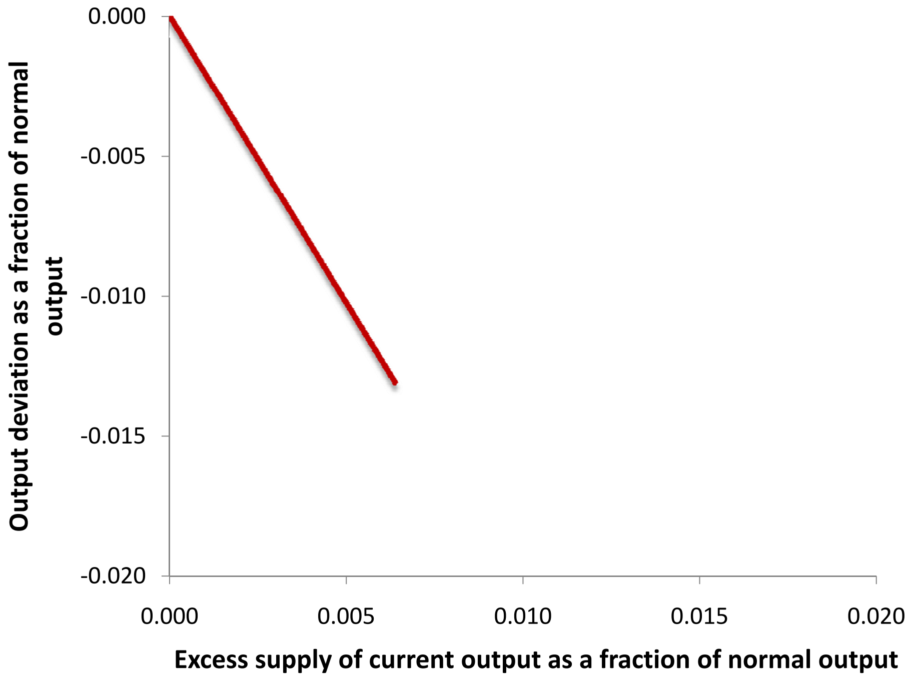

Figure 16: Relation between the Excess Supply of Goods and the Shortfall in Output

realistic assumption, but experience has shown that perfect-foresight models give surprisingly good accounts of what happens in a dynamic model once a major surprise becomes known. Thus I treat all the values of all the variables as one big vector of 792 unknowns and solve 792 nonlinear equations jointly for their exact values. I solve over a horizon of 40 years, though I show only the first 20 quarters of the solution.

8.1 Technology, adjustment costs, rental prices, and capital demand

The technology for producing output y is Cobb-Douglas with labor elasticity α:

$$y_t = n_t^{\alpha} k_{t-1}^{1-\alpha}. \tag{12}$$

Output is the production of goods, which are used to make capital, houses, and consumer durables, or are consumed directly. The price of output is $p_{y,t}$.

Capital installation occurs up to the point where the marginal adjustment cost equals the difference between the price of installed capital q_k and the price of uninstalled capital, $p_{y,t}$:

$$p_{y,t}\kappa_k \frac{k_t - k_{t-1}}{k_{t-1}} = q_{k,t} - p_{y,t}. \tag{13}$$

The parameter κ measures capital adjustment cost—if $\kappa_k = 0$, q_k is always $p_{y,t}$ and there are no adjustment costs. Housing installation follows a similar equation with subscript d replacing k.

The rental prices of capital and durables include the financial frictions $f_{k,t}$ and $f_{d,t}$:

$$p_{k,t} = (1 + r_{n,t-1})(1 + f_{k,t})(q_{k,t-1} - (1 - \delta_k)q_{k,t} \tag{14}$$

and

$$p_{d,t} = (1 + r_{n,t-1})(1 + f_{d,t})q_{k,t-1} - (1 - \delta_d)q_{d,t}. \tag{15}$$

Here $r_{n,t-1}$ is the nominal rate of interest for borrowing at the end of period $t-1$ and repaying at the end of period t.

The market-clearing condition for capital equates the marginal revenue product of capital to the rental price:

$$(1 - \alpha)\frac{p_{y,t}y_t}{k_{t-1}} = x_t p_{k,t}. \tag{16}$$

Here x_t is the markup of the product price over marginal cost.

8.2 Household product demand

Households fall into two categories, unconstrained ones who follow the standard life-cycle intertemporal model and constrained ones who are at the corner of their intertemporal choice resulting from an inability to engage in unsecured borrowing beyond a modest limit. Both types of households have active choices about the division of spending between consumption of nondurable goods and services on the one hand and the services of durable goods including housing, on the other hand. A tilde ($\tilde{}$) denotes unconstrained households and a bar ($\bar{}$) denotes constrained ones.

Consumption is a Cobb-Douglas composite of consumption of standard output, $c_{y,t}$, and the services of durables, d_{t-1}:

$$\tilde{c}_t = \tilde{c}_{y,t}^{\phi}\tilde{d}_{t-1}^{1-\phi}, \tag{17}$$

and similarly for constrained households. The price of composite consumption is

$$p_{c,t} = \phi^{-\phi}(1 - \phi)^{-(1-\phi)}p_{y,t}^{\phi}p_{d,t}^{1-\phi}. \tag{18}$$

Here $p_{d,t}$ is the rental price of durables, as above. The unconstrained household's demand for the goods component of consumption satisfies:

$$p_{y,t}\tilde{c}_{y,t} = \phi p_{c,t}\tilde{c}_t \tag{19}$$

and similarly for constrained households. Total nondurable consumption is:

$$p_{y,t}c_{y,t} = \phi p_{c,t}(\tilde{c}_t + \bar{c}_t) \tag{20}$$

Unconstrained households order their paths of composite consumption according to the intertemporal utility function

$$\sum_t \beta^t \frac{\tilde{c}_t^{1-1/\sigma}}{1-1/\sigma}, \tag{21}$$

where σ is the intertemporal elasticity of substitution.

Constrained households' consumption is

$$p_{c,t}\bar{c}_t = p_{y,t}(\omega y_t - s_t y_t), \tag{22}$$

where ω is the fraction of constrained households and s_t is the burden of interest and debt repayments of constrained households as a fraction of output.

Consumption of durables services is:

$$p_{d,t}d_{t-1} = (1-\phi)p_{c,t}(\tilde{c}_t + \bar{c}_t), \tag{23}$$

8.3 The output market and the marginal revenue product of labor

The marginal product of labor is

$$\alpha \left(\frac{k_t}{n_t}\right)^{1-\alpha}. \tag{24}$$

Firms take sales away from each other by cutting price. The marginal revenue product is less than the marginal product of labor by a factor of $1/x_t$. To sell each additional unit of output, a firm lowers its price by larger increments, so x_t is higher when rationing is more severe. When customers are not rationed and the output market clears, $x_t = 1$. When the lower bound on the interest rate binds, $x_t > 1$. In effect, x_t clears the intertemporal output market when the ZLB disables the real interest rate from playing that role. The marginal revenue product of labor,

$$m_t = \frac{p_{y,t}}{x_t}\alpha \left(\frac{k_t}{n_t}\right)^{1-\alpha}, \tag{25}$$

is the benefit to the firm from adding a worker. It is lower when customers are rationed and the markup x_t exceeds its normal value of one.

I take the price of output, $p_{y,t}$, to be an upward trend at rate π multiplied by a factor that lowers the price when firms are cutting prices in the face of customer rationing:

$$p_{y,t} = \bar{p}(1-\pi)^{-t}x_t^{-\eta}. \tag{26}$$

Because I am not offering a theory of this kind of price rigidity, I do not attempt to derive this functional form or the elasticity η from more basic foundations.

8.4 Employment

As in the standard Mortensen-Pissarides model, all workers desire to work a standard number of hours. The only source of variation in aggregate hours of work arises from unemployment.

Hall (2009b) gives a compact summary of the search-and-matching model whose canon is Mortensen and Pissarides (1994). My approach generalizes wage determination relative to the Nash bargain in that paper. Also, I simplify the treatment of labor-market dynamics by considering only the stochastic equilibrium of labor turnover, which means that the employment rate n measures the tightness of the labor market. The vacancy rate enters the picture only in fast transitional dynamics of the matching process, which can be ignored in a quarterly model without losing much. Thus the recruiting success rate is a function $q(n)$ of the employment rate. Success is higher when employment is lower.

Without loss of generality, the wage paid to the worker can be decomposed into two parts, corresponding to a two-part pricing contract (the decomposition is conceptual, not a suggestion that actual compensation practices take this form). The worker pays a present value J_t to the employer for the privilege of holding the job and then receives a flow of compensation equal to the worker's marginal revenue product.

The cost of recruiting (holding a vacancy open) is γ per period, taken to be constant in output terms. The zero-profit condition for recruiting equates the expected benefit of recruiting to its cost:

$$q(n_t)J_t = p_{y,t}\gamma. \tag{27}$$

Thus unemployment rises if J falls. I take

$$J_t = J(m_t), \tag{28}$$

an increasing function of the marginal revenue product of labor, m_t, so that, in slack markets with high x_t and thus lower m_t, a worker pays less for a job. I solve for employment as a function of m_t and take the function as constant-elastic:

$$n_t = \bar{n}\left(\frac{m_t}{\bar{m}}\right)^{\psi}, \tag{29}$$

where the elasticity ψ is positive, $\bar{m}$ is the normal level of m_t which I take to be the stationary value, and $\bar{n}$ is the normal level of employment.

The contraction in J when the marginal product of labor falls can be interpreted as wage stickiness, or, more accurately, compensation stickiness. If total compensation is sticky and the net benefit falls, then J must fall, because J is the present value of the difference between the net benefit and the worker's actual compensation.

The separation or turnover rate is a fraction ν of employment, a constant. The cost of filling a vacancy, γ/q, is 14 percent of a quarter's earnings, according to Silva and Toledo (2008).

8.5 The financial market

Only unconstrained households participate in asset markets on the margin. They price assets with returns measured in units of output by the discounter,

$$\mu_t = \beta \frac{p_{c,t}/p_{y,t}}{p_{c,t+1}/p_{y,t+1}} \left(\frac{\tilde{c}_{t+1}}{\tilde{c}_t} \right)^{-1/\sigma} . \tag{30}$$

I use the term *real interest rate* to mean the own interest rate on output. The optimal choice of consumption growth results in a discounter that discounts the market real interest rate to one:

$$(1 + r_t)\mu_t = 1. \tag{31}$$

The real and nominal interest rates are related as

$$1 + r_t = (1 + r_{n,t}) \frac{p_{y,t}}{p_{y,t+1}}. \tag{32}$$

Thus the zero lower bound on the nominal rate $r_{n,t}$ implies

$$r_t \geq \bar{r}_t = \frac{p_{y,t}}{p_{y,t+1}} - 1. \tag{33}$$

From the earlier assumption in equation (26),

$$\bar{r}_t = (1 - \pi) \left(\frac{x_{t+1}}{x_t} \right)^{\eta} - 1. \tag{34}$$

Absent the effect of customer rationing on the price level ($\eta = 0$), the bound is $\bar{r}_t = -\pi$.

8.6 Material balance

At the beginning of a period, the stock of installed capital is k_{t-1} and the stock of housing is d_{t-1}. At the end of the period, output y_t becomes available and is allocated to consumption of goods $c_{y,t}$, and investment in capital and housing, including adjustment cost, resulting in the new capital stock, k_t and new housing stock d_t. Firms expend $\gamma n_t / q(n_t)$ in recruiting cost. The equation for the economy's material balance is

$$k_t + \frac{\kappa_k}{2} \frac{(k_t - k_{t-1})^2}{k_{t-1}} + d_t + \frac{\kappa_d}{2} \frac{(d_t - d_{t-1})^2}{d_{t-1}} = (1-\delta_k) k_{t-1} + (1-\delta_d) d_{t-1} + y_t - c_{y,t} - \nu\gamma \frac{n_t}{q(n_t)}. \quad (35)$$

8.7 Customer rationing

Customer rationing occurs if the zero bound binds the interest rate, in which case firms attempt to take customers away from their rivals by cutting prices. Thus the model satisfies the complementary slackness condition: If $r_t > \bar{r}_t$, $x_t = 0$. The cases that I consider all have equilibria in which the zero bound binds for $t \in [1, T_Z]$ and not in later quarters. I search over T_Z to find the equilibrium.

8.8 Parameter values

Table 1 gives the parameter values I use in the base case and their sources. I choose a high elasticity of employment with respect to the marginal revenue product m_t to limit the volatility of the markup ratio, x_t. This choice is analogous to the high wage elasticity found necessary to rationalize the observed volatility of employment in all types of macro models. With a lower elasticity, the model would imply large movements in x_t, which would appear in the data as declines in labor's share of national income when employment falls. The cyclical stability of the share requires a high elasticity.

8.9 Initial values of endogenous state variables

Endogenous state variables are those that are updated from their initial values by amounts determined within the model. These variables are the business capital stock k and the stock of household durables including housing, d. As discussed above, I take k at its stationary value in the model and d at 1.14 times its stationary value, to incorporate the durables purchasing binge of the mid-2000s.

8.10 Exogenous variables

For the level of forced saving from tightening consumer credit standards, s, I consider a path that starts at 4.7 percent of GDP and declines to zero in equal increments over 10 quarters.

Parameter	*Description*	*Value*	*Source*
α	Labor elasticity of production function	0.646	NIPA income share
κ_k	Capital adjustment cost	8	Hall (2004)
κ_d	Durables adjustment cost	8	See text
δ_k	Capital depreciation rate	0.0188	NIPA Fixed Asset Tables
δ_d	Durables depreciation rate	0.0129	NIPA Fixed Asset Tables
ϕ	Nondurables consumption share	0.82	NIPA
β	Utility discount factor	0.9997	Derived from real federal funds rate
σ	Intertemporal elasticity of substitution	0.5	Hall (2009)
ω	Fraction of constrained consumption	0.58	See text
η	Elasticity of product price with respect to markup		See text
π	Rate of change of output price	0.01	Recent average
$\bar{n}$	Normal employment rate	0.945	Average, 1948-2007
ψ	Elasticity of employment function	10	See text
γ/q	Job-filling cost	.14 of quarterly wage	Silva-Toledo (2007)
υ	Separation rate	0.12	JOLTS, adjusted

Table 1: Parameter Values and Sources

For the financial wedges f_k and f_d, I consider a path that starts at 1.4 percent at an annual rate and declines toward zero at 5 percent per quarter ($\rho = 0.95$).

I consider two values for the rate of change of prices, π. The first is one percent per year, the rate prevailing during most of the slump through the end of 2010. The second is zero, to illustrate the effects in an economy closer to deflation.

9 Results

9.1 Characterizing the properties of a dynamic model

A dynamic model imposes a relationship on a vector of variables y_t of the form

$$y_t = F(y_{t-1}, \epsilon_t). \tag{36}$$

Here y_t is a vector of exogenous and endogenous variables and ϵ_t is a vector of random shocks. The relationship is a reduced form that derives from behavioral principles expressed in structural equations, including the life-cycle principle that calls for consumers to formulate consumption plans by looking into the indefinite future. In the limited class of models with a single representative agent and no wedges or externalities, which can be reformulated as dynamic programs, the function F comprises the policy functions and laws of motion of the dynamic program.

The sequence of derivatives

$$\frac{\partial F}{\partial \epsilon_t}, \; \frac{\partial F}{\partial y_t}\frac{\partial F}{\partial \epsilon_t}, \; \frac{\partial F}{\partial y_{t+1}}\frac{\partial F}{\partial y_t}\frac{\partial F}{\partial \epsilon_t}, \cdots \tag{37}$$

is the set of *impulse response functions* of the model. It shows how the variables in the model respond in period t and later to shocks in period t. Elements representing the response of exogenous variables to shocks in equations that determine endogenous variables are zero by definition.

The sequence of derivatives

$$\frac{\partial F}{\partial y_0}, \; \frac{\partial F}{\partial y_1}\frac{\partial F}{\partial y_0}, \; \frac{\partial F}{\partial y_2}\frac{\partial F}{\partial y_1}\frac{\partial F}{\partial y_0}, \cdots \tag{38}$$

describes how the variables respond to the initial conditions y_0. Most authors do not consider these responses and they do not have a widely used name. But the nature of the current investigation gives them an important role. I will call them *initial condition response functions.*

I noted earlier that the model does not treat uncertainty explicitly, even though it is obvious that households and producers lack anything like perfect foresight. I exploit the

Law of Dynamic Responses

The initial condition response functions of the perfect-foresight version of a model are surprisingly good estimates of the impulse response functions of a stochastic version of the same model, where the perfect-foresight version is solved as if the realization of a random shock had just entered the initial conditions of the model.

The law is on the same footing as Krusell and Smith's (1998) finding that a small number of well-chosen moments can stand in for an entire distribution in a dynamic general-equilibrium model. It is less than a universal truth.

9.2 High initial stock of durables

The first dynamic equilibrium I calculate is for an economy that begins with a stock of durables 14 percent above the stationary value. No households are liquidity-constrained in this economy and there are no financial frictions. The trend rate of inflation, π, is one percent per year at all times, or, in the variant, zero. Thus the economy's lower bound on the safe short-term real rate is either about minus one percent per year (I say "about" because of the effect of changes in the markup on inflation when the bound binds) or zero.

Table 2 gives the results for this and other exercises using the model and also reports, in the bottom row, data for the current slump. The left panel shows the decline in output, averaged over the first four quarters in the case of the model and for two years into the slump in the case of the actual U.S. economy. The right panel shows the annual real interest rate in basis points.

The economy hit only by the elevated level of durables and with positive inflation has adequate headroom so that the real interest rate can clear the current market for output without encountering the lower bound. The left column in each panel describes this case. The immediate level of the annual real rate is minus 9 basis points compared to a normal level of plus 12 basis points. Durables investment starts well below its normal level, but

	Employment shortfall, percentage points		*Annual real interest rate, basis points*	
	1 percent inflation	*Zero inflation*	*1 percent inflation*	*Zero inflation*
Durables elevation only	0.0	2.1	-9	0
Consumer credit tigntening only	1.7	6.7	-100	0
Financial friction only	0.0	18.5	-100	0
All three together	10.5	48.4	-100	0
Actual	4.5		-77	

Table 2: Responses of Employment and the Real Interest Rate

the lower real rate results in a fully offsetting increase in investment in business capital. Essentially nothing happens to consumption of nondurables and services.

The situation is altogether different if the economy has less headroom, as shown in the right column in each panel, for zero inflation. This economy cruises at full employment in its stationary state, but is vulnerable to small disturbances. In particular, if it starts with the same 14 percent bulge in its durables stock, employment is 6.7 percent below normal. In this case, the real interest rate is constrained at zero by the zero inflation rate.

9.3 Overhang of consumer debt service obligations

The second line in Table 2 shows the effects of forced saving from a tightening of consumer credit. Even with one percent per year inflation, the adverse effects push the economy to the region of the lower bound, so employment is 1.7 percentage points below normal. The real interest rate is at the bound of minus 100 basis points. If the rate of inflation is zero, the bound binds much more influentially. Employment is 6.7 percentage points below normal. The interest rate is at the bound of zero real.

9.4 Financial friction

The third line of Table 2 shows that the economy with one percent annual inflation has enough headroom to handle the friction of 140 basis points at annual rates. The headroom is only 112 basis points (from a normal real rate of plus 12 basis points to a constrained rate of minus 100 basis points), but the friction does not depress the real rate earned by savers by the full amount of its 140 basis points. About half of the effect is to raise the rate paid by borrowers. See Hall (2011) for further discussion of this point.

In the economy with zero inflation, the financial friction causes a calamity—employment is 18.5 percentage points below normal.

9.5 Combined effects of the three forces

The fourth line of Table 2 describes the effects of all three adverse effects combined. With one percent inflation, employment is 10.5 percentage points below normal, about twice as deep a slump as actually occurred. With zero inflation, the combined effect is worse than the Great Depression, with employment just over half its normal level.

9.6 Conclusions from the model

These results demonstrate the potential extreme sensitivity of economic activity in a low-inflation economy whose real interest rate has inadequate room to decline to offset a force that lowers the equilibrium real rate. The very same economy without a lower limit on the real rate is completely stable and always operates at full employment.

The model shares a feature with almost all dynamic general-equilibrium models—it has almost no ability to generate long slumps endogenously. The persistence of the driving forces controls the persistence of low employment. In the case of the overhang of housing and consumer durables, the model generates only moderate persistence. If that were the only factor explaining the slump, the economy would be well on the way to recovery as of the end of 2010.

Evidence on the persistence of the two credit-related driving forces—consumer credit standards and the financial friction—is mixed. This topic needs a lot more investigation.

10 Simple Sticky-Price and Sticky-Wage Models

The model developed here has three important properties that are departures from standard economic thinking about prices and wages: (1) the rate of inflation in the price of output is close to exogenous , (2) the markup ratio rises as customer rationing becomes more intensive, and (3) real wages are sticky. Various forms of these departures have been found necessary in a number of modern dynamic general-equilibrium macro models. In this section, I investigate whether any simpler specifications, closer to standard economic principles, might deliver reasonable accounts of the facts about persistent slumps.

One possibility is to drop property (3), sticky real wages, in favor of normal equilibrium in a competitive labor market without search frictions. In the model, this alternative corresponds to setting the parameter ψ, the elasticity of employment with respect to the markup ratio, to zero. Equilibrium in the resulting model has a markup ratio less than one, which is impossible. With the added assumption that the ratio is one—that product markets are always competitive—the model has no equilibrium. In effect, the model with these two modifications is a standard real business cycle model, which determines the real interest rate. Forcing the interest rate upward by imposing the zero lower bound places the model out of equilibrium. A standard solution to this problem is to drop labor-market equilibrium and require intertemporal equilibrium in the output, which amounts to assuming a sticky real wage—it generates an implicit flat labor supply relation.

Another possibility is to drop properties (1) and (2), and instead to take the output market to be perfectly competitive and always in equilibrium. The simplest version of the model assumes that labor supply is perfectly elastic at a fixed real wage. Shimer (2010b) takes this approach. The resulting model generates persistent slumps if some force causes a reduction in the full-employment marginal product of labor. A fall in productivity would have that effect. Because productivity rose at unusually high rates during the contraction phase of the current slump, explanations of the slump based on declining productivity are not plausible.

Shimer proposes a fall in the business capital stock as a driving force. Because that event drives down the full-employment marginal product of labor, the loss of capital causes an immediate contraction. The level of employment falls by the same proportion as the capital stock, thus restoring the marginal product of labor to the level of the rigid real wage. A slump ensues until capital accumulation and productivity growth raise the full-employment marginal product of capital back to the level of the fixed real wage.

Data on business capital show no decline around 2008. Thus Shimer's model in its stated form does not explain the current slump unless there has been a decline in the effective capital stock not captured in the standard data. At the end of the paper he suggests one approach that might have this implication.

Shimer's model differs from the one developed here mainly in its lack of concern with the decline of the safe short-term nominal interest rate to zero during the current U.S. slump and in other slumps, such as the one that began 20 years ago in Japan. An extension of his model to this topic would require adding features that deal with the nominal interest rate. The assumption of a competitive output market, together with almost any standard specification of the determination of the price level, would probably imply that the zero lower bound on the nominal interest rate does not matter. The reason is that, in flexible-price monetary models, the immediate effect of an adverse shock is a discontinuous drop in the price level followed by a gradual rise. The expected rate of inflation rises the moment the shock hits. Thus the nominal interest rate rises and the lower bound of zero is irrelevant.

Adding sticky inflation to Shimer's model would also require the model to take a stand on what happens when the intertemporal output market has excess current supply. The simplest answer—embodied in Krugman (1998) and Christiano et al. (2009)—is that the quantity transacted is the amount demanded. When combined with the assumption of a fixed real wage, this results in a model that accounts nicely for slumps. It is a fairly close cousin of the model developed here, which differs only by granting sellers of output some power to take sales away from rivals when customers are rationed and considers how this factor enters a DMP-style model of the labor market.

11 Restructuring in Slumps

The model developed here has the property that full employment will resume as soon as real asset returns move to the point where there is no longer an excess supply of current output. In Friedman's terminology, the natural rate of unemployment is an unchanging constant. Deviations around the natural rate arise from sources outside the labor market.

Many economists in this and earlier slumps have attributed substantial amounts of the rise in unemployment in slumps to increases in the natural rate. The increase is often associated with the concept of *structural unemployment*, though that concept has not emerged as a distinct type of unemployment in the modern turnover-based theory of unemployment. Two specific sources of a rise in the natural rate have been the focus of controversy in the current slump: (1) the increased subsidy to unemployment from the extension of unemployment benefits to 99 weeks, and (2) the collapse of employment in construction and some manufacturing industries, which created a huge mismatch between job-seekers—blue-collar men—and job openings in health, other services, and in industries calling for knowledge

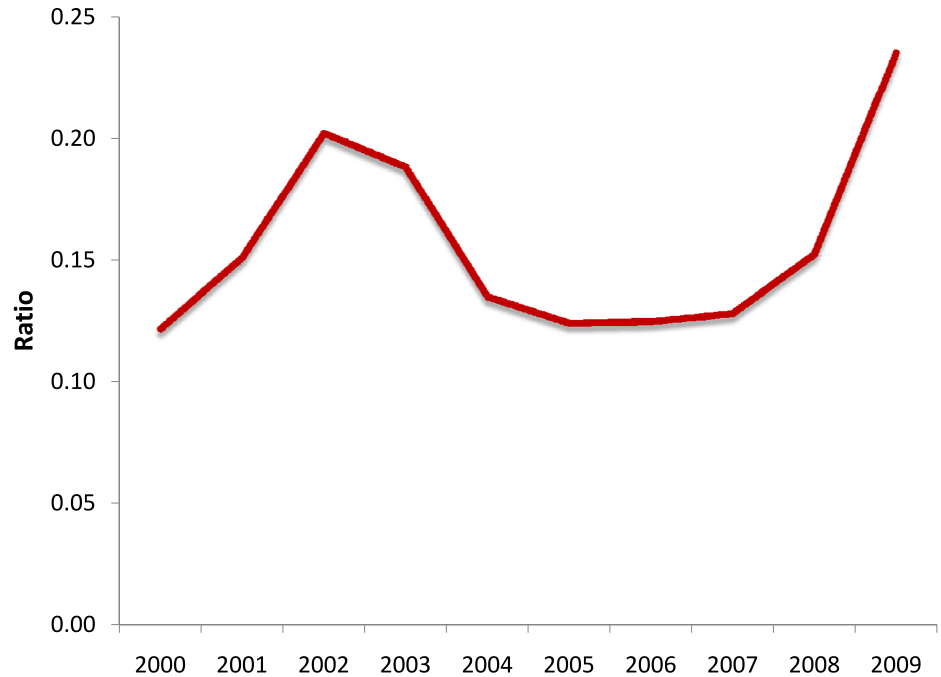

Figure 17: Ratio of UI Benefits to Median Earnings, 2000 to 2009

workers.

11.1 Unemployment benefits

One of the forces tending toward higher unemployment is the increase in unemployment benefits, with the potential to last for almost two years past job loss, well beyond the duration of extensions in earlier recessions. Figure 17 shows an estimate of the effective replacement rate for benefits in the past decade. It is the ratio of total benefits as reported in the National Income and Product Accounts, Table 3.12, to the median weekly earnings for full-time workers from the Current Population Survey, multiplied by 52. The UI replacement ratio is somewhat higher than in the previous recession, but by an amount that is surprisingly small given how much higher unemployment has risen in this one. Further, the level of the ratio, less than 25 percent of median earnings, does not seem high enough to have a large effect. Still, UI benefits are a subsidy to unemployment and it seems likely that the general principle holds that a subsidy induces more of the subsidized behavior.

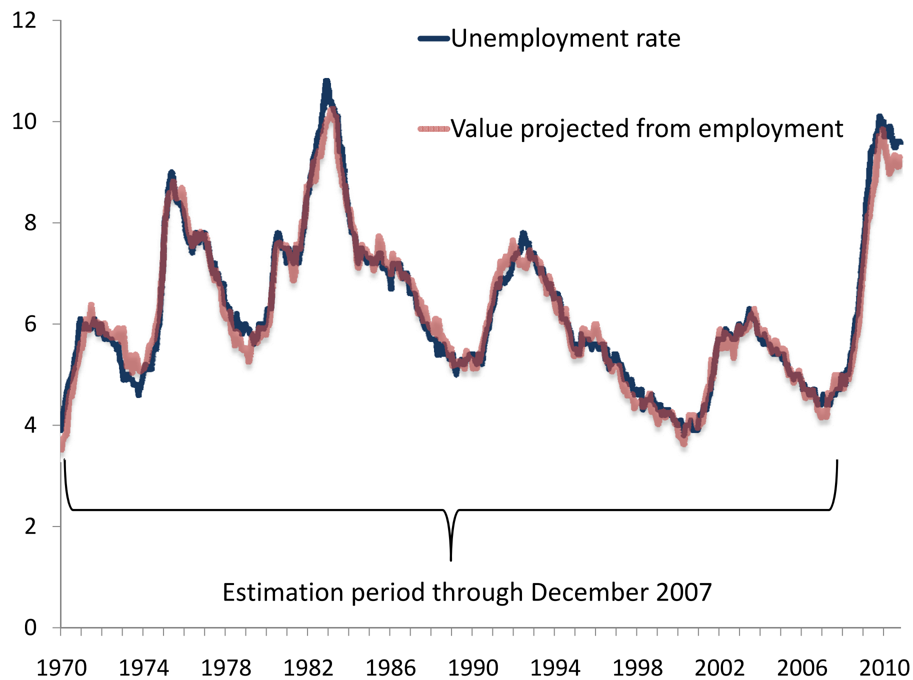

Figure 18: Actual Unemployment Rate and Rate Projected from Historical Relation between Unemployment and Employment

11.2 Shift in unemployment's relation to employment

One of the ways that the increased unemployment subsidy may increase unemployment is to induce people to remain in the labor force as insured job-seekers rather than dropping out of the labor force after losing jobs. The result would be higher unemployment than would otherwise occur. Alternatively, the increased subsidy might result in job-seekers turning down job offers that they would have otherwise accepted, resulting in lower employment for given levels of unemployment. To generate a plausible counterfactual path for unemployment, I exploit the stable and close relationship between the employment and unemployment rates that prevailed prior to the current slump.

Figure 18 shows the actual unemployment rate since 1970 and the rate projected from a regression where the right-hand variables are the contemporary employment/population ratio, that ratio lagged 12, 24, and 36 months, and the first through fourth powers of a time trend, using data through December 2007. The out-of-sample projection during the slump shows that the decline in the employment/population ratio accounts for all but a small part of the explosion of unemployment. Conditional on the severe decline in the employment/population ratio, the big increase in unemployment is not a surprise.

I conclude that the joint behavior of the labor force, unemployment, and employment, was almost the same in the current slump as in past experiences, including in particular the previous serious contraction starting in 1981. There is essentially no sign of excess unemployment from extended unemployment benefits or other special forces in the years after 2007. In particular, it casts doubt on the hypothesis that mayhem in the housing market reduced geographic mobility that might otherwise have helped limit unemployment.

The rise in unemployment starting at the end of 2007 was smaller as a fraction of the population than was the decline in the employment/population ratio. That is, the labor force as a fraction of the population declined substantially during the contraction of 2008 and 2009. Because the labor force is just the sum of unemployment and employment, the finding of no surprise in the joint behavior of unemployment and employment automatically implies that there was no surprise in the behavior of the labor force during the contraction.

Notice that these statements do not require me to take a stand on the forces leading to fluctuations in employment and unemployment. Though I am inclined to think that factors outside the labor market caused a reduction in employment and that rise in unemployment and the decline in the labor force were normal reactions to that large reduction, the findings are equally relevant within a line of thought where changes in the labor force cause changes in employment. In that view, a large reduction in the labor force resulted in an even larger decline in employment and a corresponding increase in unemployment.

11.3 Reconversion after World War II

The mismatch-restructuring view holds that it takes several years at best for the labor market to absorb workers released from a collapsing sector into employment in other sectors. The U.S. put this proposition to test in the second half of 1945 and in 1946, when 8 million members of the armed forces returned to civilian life. Table 3 shows what happened. 7.8 million men returned to normal activities. Only 240,000 left the labor market for reasons other than going to school. Employment of men rose by 4.7 million or 10 percent of the male labor force as of 1945. Unemployment among men rose by 2.6 percent of the labor force, bringing the male unemployment rate to 4.4 percent in 1946. The increase in employment was 80 percent of the increase in the labor force. All this happened in a single year.

The decline in employment of women was half of the growth in employment for men. Women were drawn into the work force during the war. Whatever adjustments occurred

	Men		Women	
	Number (thousands)	*Percent of male labor force*	*Number (thousands)*	*Percent of female labor force*
In armed forces	-7,800	-17.0	-180	-0.9
Civilian employment	4,730	10.3	-2,300	-11.9
Unemployment	1,180	2.6	77	0.4
School	2,070	4.5	390	2.0
Not in labor force, other activities	240	0.5	2,620	13.6
Total (population growth)	420	0.9	607	3.1
1946 unemployment rate		4.4		3.3

Table 3: Changes in Activities of the U.S. Population between 1945 and 1946

during the war to accommodate a large increase in employment of women were reversed quickly in 1946, when 2.6 million women or 13.6 percent of the 1945 female labor force left the labor force.

The point of studying data for the unusual events in the immediate aftermath of World War II is to demonstrate that there is no mechanical obstacle to a huge rearrangement of work activities. Incentives to create jobs and recruit workers were strong. In contrast to the situation in 2008, when the U.S. economy had an overhang of houses and consumer durables from the spending spree earlier in the decade, in 1946 Americans were starved for durable goods. Monetary policy kept the short-term safe rate close to zero.

The 8 million returning members of the armed forces were 15 percent of the level of civilian employment in 1945. By contrast, the decline in employment in the sector most obviously a source of mismatch in the current slump, construction, from December 2007 to December 2009, was 1.3 percent of total employment.

There is no possibility of recreating today the special circumstances that made reconversion so successful in 1945 and 1946. The lesson of that experience is that the labor market can handle a large volume of potential mismatch in a single year if incentives for job creation are strong enough.

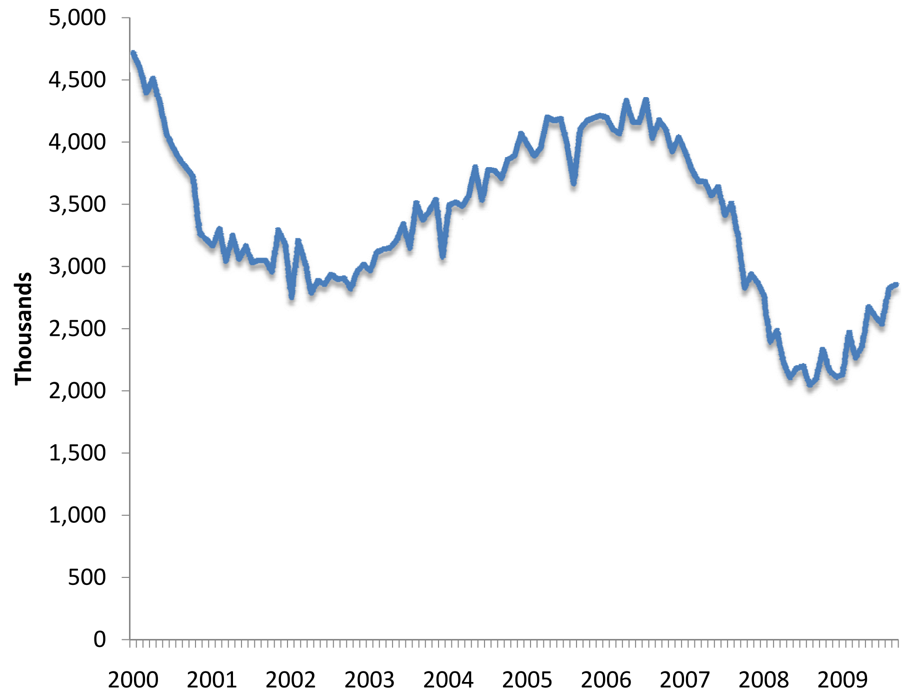

Figure 19: Number of Job Openings, 2000 to 2010

11.4 Job openings and the matching process

Figure 19 shows the number of job openings as measured in JOLTS, the Job Openings and Labor Turnover Survey. Openings are highly cyclical. In a slack labor market, jobs fill much faster, so the total number of openings is low because only those recently posted remain unfilled. Openings reached a trough in July 2009 and have risen smartly since then, while employment continued to decline until December 2009 and unemployment is only slightly below its peak. Some observers have wondered what factors accounted for the difference in the behavior of openings and unemployment and even speculated that openings might be a better measure of the state of the labor market in the current slump because some factor—again, generous UI benefits—was raising unemployment artificially.

Pissarides (1987) introduced the matching function to unemployment theory. It provides a natural framework for measuring matching efficiency. The flow of job matches or hires H is a function of the number of searching workers U and the number of job openings or vacancies, V. A large body of empirical research surveyed in Petrongolo and Pissarides (2001) suggests that

$$H_t = \phi_t \sqrt{U_t V_t} \tag{39}$$

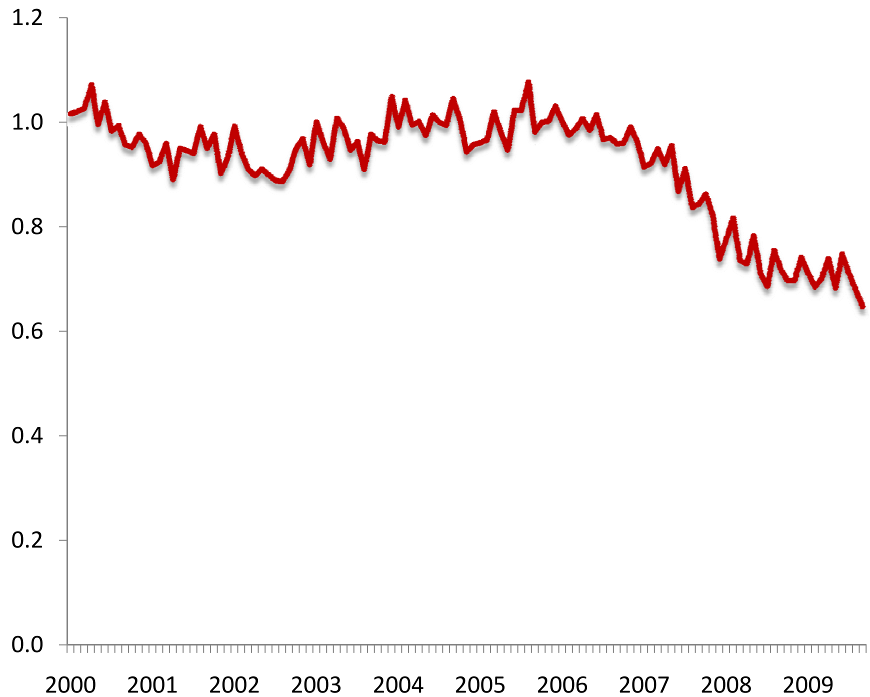

Figure 20: Efficiency of Matching Job-Seekers to Jobs

is a good representation of the matching function. Here ϕ_t is the efficiency of matching. Figure 20 shows the efficiency calculated as

$$\phi_t = \frac{H_t}{\sqrt{U_t V_t}} \tag{40}$$

from JOLTS data on hires H_t and openings V_t along with the household survey data on the number of unemployed workers U_t.

Matching efficiency is pro-cyclical. It fell in the slump that started in early 2001, rose during the mid-decade boom, then fell sharply in the severe recession. Just as it lagged behind the recovery in 2002, it has not yet turned up, despite modest job growth and noticeable growth in job openings.

The Beveridge curve is a convenient way to look at the joint behavior of unemployment and job openings. Figure 21 plots the unemployment rate on the horizontal axis and the ratio of JOLTS job openings to the labor force on the vertical axis (this is not exactly the rate published in JOLTS, which is the ratio of openings to employment). The blue diamonds are for the months December 2000 through July 2009, when the curve was reasonably stable. The last 13 months showed a distinct upward shift, corresponding to the drop in matching efficiency. Job openings increased substantially but the usual decline in unemployment did

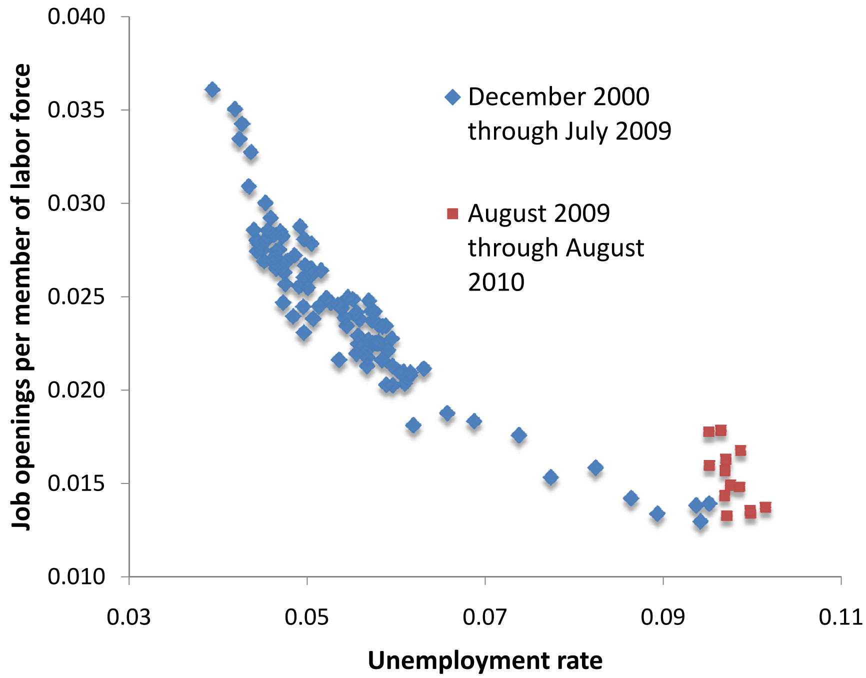

Figure 21: The Beveridge Curve, 2000 through 2010

not occur. Many commentators have pointed to this development as another indication that unemployment was above its expected level.

The Beveridge curve summarizes changes in the labor market's equilibrium over the business cycle. No principle underlying that equilibrium guarantees the stability of the curve—quite the contrary, there is a large literature on loops in the curve over the cycle. The matching function helps explain the issues in shifts of the Beveridge curve. By dividing both sides of the matching function in equation (39) by the labor force L_t and rearranging, one gets

$$\frac{U_t}{L_t} = \left(\frac{H_t/L_t}{\phi_t}\right)^2 \frac{L_t}{V_t} \tag{41}$$

or

$$u_t = \left(\frac{h_t}{\phi_t}\right)^2 \frac{1}{v_t}, \tag{42}$$

where lower-case letters denote ratios to the labor force. The bottom line is that unemployment will be proportional to the reciprocal of the job openings rate if the ratio of the hiring rate to match efficiency, h_t/ϕ_t, is constant. Another way to express this proposition is that the Beveridge curve will be a hyperbola if that ratio is constant. An increase in the hiring rate h_t or a decrease in matching efficiency ϕ_t will shift the Beveridge curve outward. The reason for the stability of the Beveridge curve prior to August 2009 was that the hiring rate and matching efficiency tended to move together quite closely.

The traditional view of the Beveridge curve is that the labor market makes counterclockwise loops in the (u, v) space. The rationalization for this view is that matching efficiency is roughly constant, but the hiring rate h_t is low during a contraction, as employment falls, and then is high in the recovery, as employment rises. The result is that the factor $\left(\frac{h_t}{\phi_t}\right)^2$ is low in the contraction phase and high in the expansion phase, resulting in the clockwise loops. In the U.S., which has had economy-wide official data on job openings only since the beginning of the 2001 recession, the loops have not previously been apparent. Hiring was only slightly pro-cyclical—constancy of turnover was not a bad approximation—and what variations in hiring that did occur were offset by comparable variations in matching efficiency.

The sharp decline in matching efficiency that accompanied the recession that began at the end of 2007 was not the only cause of the outward shift of the Beveridge curve. An outward loop would have occurred on account of the severity of the recession even with normal behavior of matching efficiency. To illustrate this point, I consider a counterfactual scenario for the labor market in the recession with the following features: (1) the hiring rate would have been the same as actually occurred but (2) matching efficiency would be the equally weighted average of its actual value and its value in April 2008, just before the decline in matching efficiency began. In this scenario, the decline in efficiency is about double the decline that occurred in the aftermath of the 2001 recession, which strikes me as a reasonable criterion for normal behavior in the labor market.

Figure 22 shows the counterfactual Beveridge curve in the same format as Figure 21. A prominent loop is apparent. Procyclical variations in the hiring rate were a significant factor in the behavior of the labor market. From this finding, I conclude that the adverse shift of the Beveridge curve is in part a normal response of the labor market and not entirely the result of some special force tending toward higher unemployment.

11.5 Sources of the decline in matching efficiency

The fact remains, of course, that unemployment is much higher than it would have been absent the dramatic decline in matching efficiency. Research on the severity of unemployment should concentrate on this important fact.

Here I will discuss one factor that points in the direction of lower efficiency during slumps,

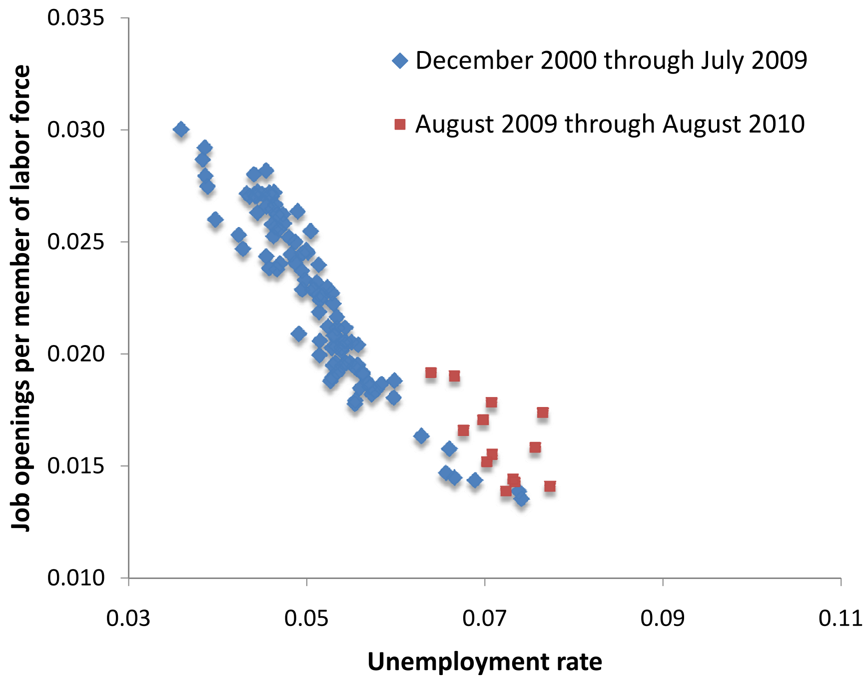

Figure 22: Counterfactual Beveridge Curve with Normal Behavior of Matching Efficiency

based on the lower rate of turnover that results from the low level of quits when jobs are hard to find. The basic idea is that it is easy for employers to recruit replacements for quitting workers—they tend to have short tenure and low and generic skills. Though such workers are a small fraction of employment, they are a much higher fraction of hiring and separation flows.

Figure 23 shows two measures of labor turnover on the hiring side each calendar quarter. The top line shows data from JOLTS, which asks employers how many workers were hired each month. I have added together data for the three months of each quarter. JOLTS understates hires because it does not consider newly opened establishments. The lower line shows a different measure from the Business Employment Dynamics data, also compiled by the Bureau of Labor Statistics. The BED uses administrative data covering essentially all businesses. The series—often called *job creation*—measures increases in all units that had gains in employment. This measure understates actual hiring because some units hire workers in the same quarter that they lay others off. Davis, Faberman and Haltiwanger (2010) shows how this overlap occurs. Their Figure 6 reports the rates of hiring among shrinking firms and the rates of separations among expanding ones. They are in the process of preparing estimates of hiring and other flows from the gross employment changes in BED.

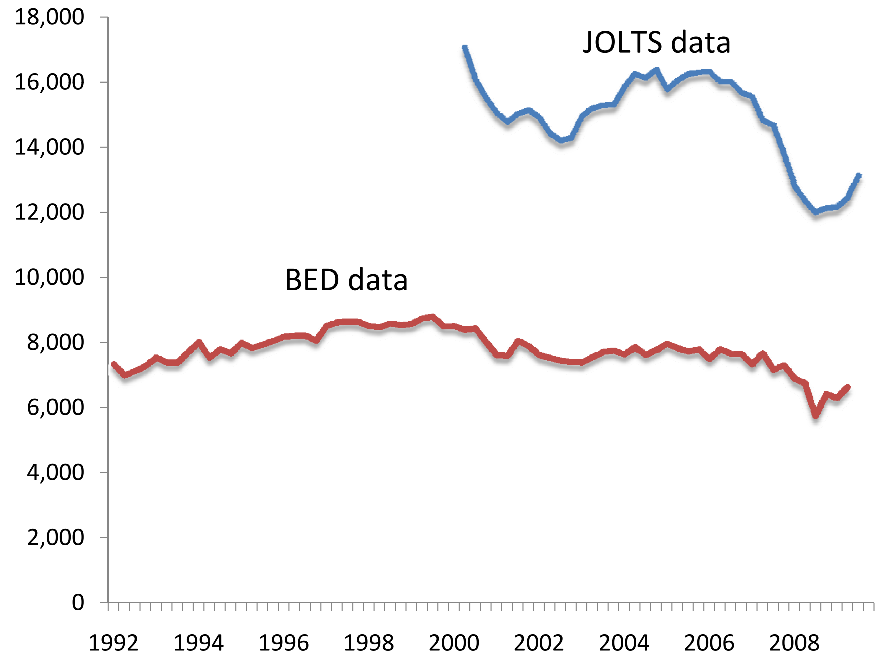

Figure 23: Two Measures of Numbers of New Hires per Quarter

Hiring rates in JOLTS are about twice as high as job-creation rates in BED and are also more variable. My understanding is that Davis et al.'s research is able to reconcile both of these dimensions of the behavior of the two bodies of data. In times when jobs are easy to find, quits are more frequent but layoffs are at normal levels, so turnover is higher. Thus the ratio of the JOLTS hiring rate to the BED job-creation rate is a measure of turnover. Figure 24 shows the high correlation between the flow of hires and quits in JOLTS.

In strong labor markets, employers need to replace high-turnover quitting workers in disproportionate numbers. Because workers who have quit at other firms are widely available, the matching process is easy and efficient. On the other hand, when jobs are harder to find, quit-related turnover is much lower and the recruiting process is focused more on creating long-term matches. The process is more time-consuming and appears to be less efficient, though in fact it is solving a harder problem.

The decline in match efficiency is closely related to the upward shift in job openings relative to the number of unemployed workers. My interpretation emphasizes that a lower fraction of high-turnover job filling will lower measured efficiency. Others believe that the higher fraction of job-seekers who have been out of work for a year or more reduces matching efficiency because of deteriorating job skills.

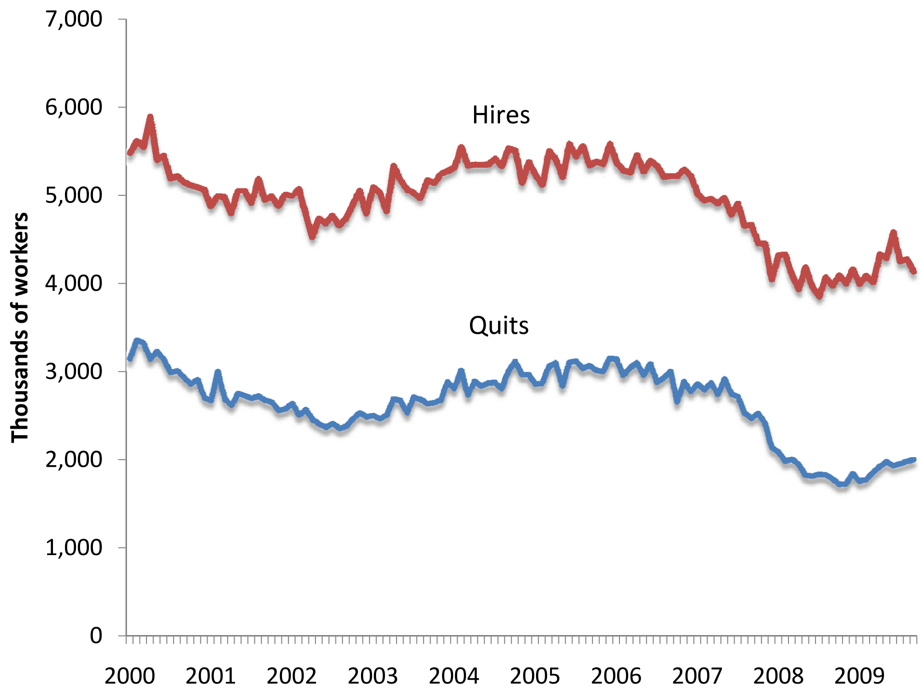

Figure 24: Numbers of Workers Hired and Numbers of Quits, JOLTS

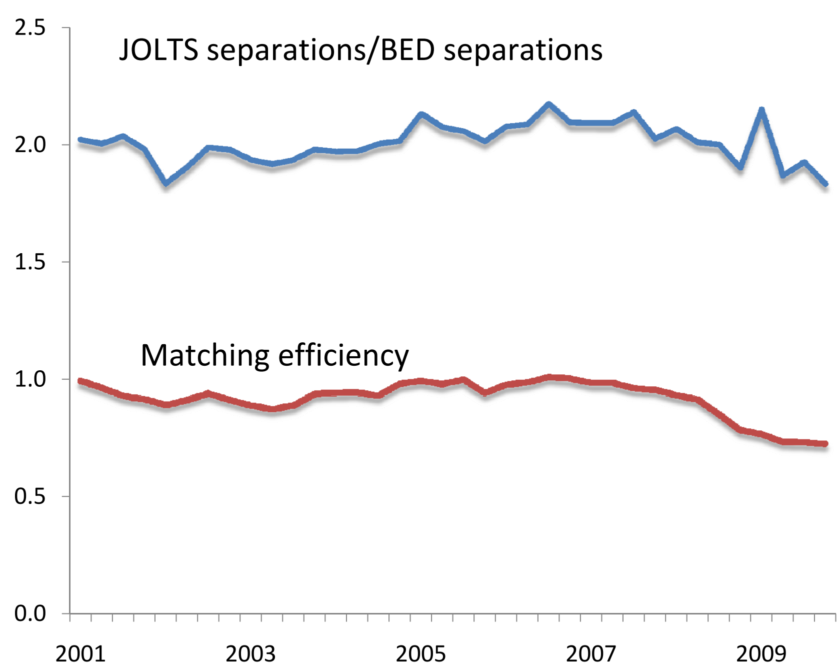

Figure 25: Ratio of JOLTS Hiring Rate and BED Job-Creation Rate together with Matching Efficiency

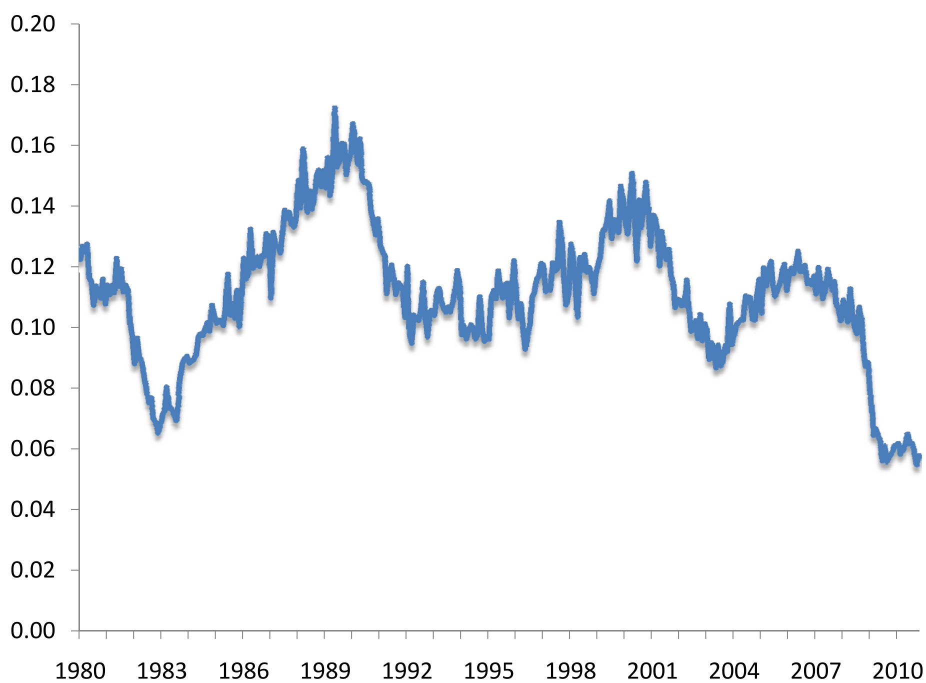

Figure 26: Fraction of Unemployed Who Quit Previous Jobs

Figure 26 shows the fraction of the unemployed who have quit jobs and confirms the sharp decline of the fraction during slumps.

11.6 Conclusions on mismatch and structural unemployment

Mismatch can be an important problem in the labor market, as it was at the end of 1945. But the U.S. labor market is capable of managing large volumes of mismatch without high or prolonged unemployment. When job-creation incentives are weak, as they are in slumps, the mismatched jobless, such as construction workers today, suffer high unemployment, along with many other job-seekers with readily marketable skills.

The distinction between cyclical and structural unemployment is unproductive. Structural unemployment has melted away in almost every expansion, including the most recent one in the 2000s. Still, there is much more to learn about persistent unemployment in slumps, especially about the decline in matching efficiency in the two slumps that have occurred since the U.S. began measuring the concept accurately.

12 Concluding Remarks

An economy with a disabled real interest rate is in deep trouble when one type of spending—homebuilding and consumer durables in the current slump—declines. A slump will last until the affected spending resumes its normal level. Consequently, the slump may last many years.

The analysis and calculations in this paper assume that the gradual price adjustment described by the Phillips curve does not occur. Inflation remains at the same rate. If inflation declines and turns into growing deflation, the slump will worsen, as the real interest rate rises. So far in the current slump, notwithstanding episodes of grave concern, no slide into deflation has occurred.

Absent a radical and highly unlikely monetary reform that permits negative nominal interest rates, the types of policies that could ameliorate the slump are those that emulate the effect of low real rates—making current purchasing cheaper than future. A key feature of these policies is to defer the time when the policy reverses itself until after full employment prevails. The cash-for-clunkers program in 2009 induced a significant bulge in car purchases, but because it lasted only a few months, it only deferred purchases for that many months and did nothing to shift purchases from a time of full employment to the present (Mian and Sufi (2010)). The effective program would place a high subsidy on current purchasing and phase out the subsidy, eventually becoming a consumption tax that financed the earlier subsidy. The shrinkage rate of the subsidy would amount to a negative real interest rate in consumer purchasing decisions.

References

Buiter, Willem H., "Negative nominal interest rates: Three ways to overcome the zero lower bound," *The North American Journal of Economics and Finance*, 2009, *20* (3), 213 – 238.

Burnside, Craig, Martin Eichenbaum, and Sergio Rebelo, "Booms and Busts: Understanding Housing Market Dynamics," June 2010. in preparation.

Christiano, Lawrence J., Cosmin Ilut, Roberto Motto, and Massimo Rostagno, "Monetary Policy and Stock Market Boom-Bust Cycles," March 2007.

——, Martin Eichenbaum, and Charles L. Evans, "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy," *Jounal of Political Economy*, 2005, *113* (1), 1–45.

——, ——, and Sergio Rebelo, "When Is the Government Spending Multiplier Large?," October 2009. NBER Working Paper 15394.

——, Mathias Trabandt, and Karl Walentin, "Introducing Financial Frictions and Unemployment into a Small Open Economy Model," April 2010.

Davis, Steven J., R. Jason Faberman, and John Haltiwanger, "Labor Market Flows in the Cross Section and Over Time," September 2010. Chicago Booth School of Business.

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

Gertler, Mark and Peter Karadi, "A Model of Unconventional Monetary Policy," April 2009.

Hall, Robert E., "Measuring Factor Adjustment Costs," *Quarterly Journal of Economics*, August 2004, *119* (3), 899–927.

——, "By How Much Does GDP Rise If the Government Buys More Output?," *Brookings Papers on Economic Activity*, 2009, (2), 183–231.

——, "Reconciling Cyclical Movements in the Marginal Value of Time and the Marginal Product of Labor," *Journal of Political Economy*, April 2009, *117* (2), 281–323.

——, "Why Does the Economy Fall to Pieces after a Financial Crisis?," *Journal of Economic Perspectives*, Fall 2010, *24* (4), 3–20.

——, "The High Sensitivity of Economic Activity to Financial Frictions," *Economic Journal*, 2011. Forthcoming.

Krugman, Paul R., "It's Baaack: Japan's Slump and the Return of the Liquidity Trap," *Brookings Papers on Economic Activity*, September 1998, (2), 137–205.

Krusell, Per and Jr. Smith Anthony A., "Income and Wealth Heterogeneity in the Macroeconomy," *Journal of Political Economy*, 1998, *106* (5), pp. 867–896.

Mian, Atif and Amir Sufi, "Household Leverage and the Recession of 200709," *IMF Economic Review*, July 2010, *58* (2), 74–117.

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

Petrongolo, Barbara and Christopher A. Pissarides, "Looking into the Black Box: A Survey of the Matching Function," *Journal of Economic Literature*, 2001, *39* (2), pp. 390–431.

Pissarides, Christoper A., "Search, Wage Bargains and Cycles," *Review of Economic Studies*, 1987, *54*, pp. 473–483.

Shimer, Robert, "The Cyclical Behavior of Equilibrium Unemployment and Vacancies," *American Economic Review*, 2005, *95* (1), 24–49.

——, *Labor Markets and Business Cycles* CREI Lectures in Macroeconomics, Princeton University Press, 2010.

——, "Wage Rigidities and Jobless Recoveries," November 2010. Department of Economics, University of Chicago.

Silva, Jose A. and Manuel Toledo, "Labor Turnover Costs and the Cyclical Behavior of Vacancies and Unemployment," February 2008. Universitat Autonoma de Barcelona and Universidad Carlos III de Madrid.

Stock, James H. and Mark W. Watson, "Modeling Inflation after the Crisis," *Economic Policy Symposium, Federal Reserve Bank of Kansas City*, 2010. Forthcoming.

Townsend, Robert M., "Optimal contracts and competitive markets with costly state verification," *Journal of Economic Theory*, 1979, *21* (2), 265 – 293.

Appendix

A Constructing the Index of Lending Standards

The model of bank i's change in standards is

$$\delta_i = \Delta x + \mu + \epsilon_i, \tag{43}$$

where ϵ_i has the standard normal distribution. Then

$$\begin{aligned} \Pr[\text{Bank } i \text{ eased}] &= \Pr[\delta_i < -\nu] & (44) \\ &= \Pr[\Delta x + \mu + \epsilon_i < -\nu] & (45) \\ &= \Pr[\epsilon_i < -\Delta x - \mu - \nu] & (46) \\ &= \Phi(-\Delta x - \mu - \nu) & (47) \end{aligned}$$

and, by similar logic,

$$\Pr[\text{Bank } i \text{ tightened}] = \Phi(\Delta x + \mu - \nu). \tag{48}$$

Thus

$$\begin{aligned} \text{Net change} &= \Pr[\text{Bank } i \text{ tightened}] - \Pr[\text{Bank } i \text{ eased}] & (49) \\ &= \Phi(\Delta x + \mu - \nu) - \Phi(-\Delta x - \mu - \nu) & (50) \end{aligned}$$

B Model Details

B.1 Stationary model

Auxiliary:

$$r = \frac{1}{\beta} - 1 \tag{51}$$

$$n = \bar{n} \tag{52}$$

$$p_k = r + \delta_k \tag{53}$$

$$p_d = r + \delta_d \tag{54}$$

$$c_y = p_c c \phi \tag{55}$$

Core:

$$y = n^{\alpha} k^{1-\alpha} \tag{56}$$

$$k = (1-\alpha)\frac{y}{p_k} \tag{57}$$

$$p_d d = (1-\phi) p_c c \tag{58}$$

$$y = c_y + \delta_k k + \delta_d d + \gamma \nu \bar{n} \tag{59}$$

$$p_c = \phi^{-\phi}(1-\phi)^{-(1-\phi)} p d^{1-\phi} \tag{60}$$

B.2 Dynamic model

Timing: Output in t uses workers n_t and end-of-period capital k_{t-1}. Consumption occurs at the end of t and comes out of output at the end of t, y_t. The real interest rate from t to $t+1$ is r_t. The rental price for capital used during t is $p_{k,t}$. Capital in use during period t occurs during the period governed by r_{t-1}. That capital was acquired in $t-1$ at price $q_{r,t-1}$. The stochastic discounter from t to $t+1$ is μ_t. The ZLB binds in periods 2 through T_Z. All prices are restated here are restated in output units; they correspond to prices divided by $p_{y,t}$ in the body of the paper.

Core variables: p_d, $\tilde{c}$, n, k, d, and r.

Auxiliary:

$$n: \; n_t = \bar{n}, t \in [T_Z + 1, T] \tag{61}$$

$$y: \; y_t = n_t^{\alpha} k_{t-1}^{1-\alpha}, t \in [2, T] \tag{62}$$

$$m: \; m_t = \bar{m} \left(\frac{n_t}{\bar{n}} \right)^{1/\psi}, t \in [2, T] \tag{63}$$

$$x: \; x_t = \frac{\alpha}{m_t} \left(\frac{k_{t-1}}{n_t} \right)^{1-\alpha}, t \in [2, T_Z] \tag{64}$$

$$x: \; x_t = 1, t \in [T_Z + 1, T] \tag{65}$$

$$p_k: \; p_{k,t} = \frac{1-\alpha}{x_t} \left(\frac{k_{t-1}}{n_t} \right)^{-\alpha}, t \in [2, T] \tag{66}$$

$$p_c: \; p_{c,t} = \phi^{-\phi} (1-\phi)^{-(1-\phi)} p_{d,t}^{1-\phi}, t \in [2, T] \tag{67}$$

$$\mu: \; \mu_t = \beta \frac{p_{c,t}}{p_{c,t+1}} \left(\frac{\tilde{c}_{t+1}}{\tilde{c}_t} \right)^{-1/\sigma}, t \in [2, T-1] \tag{68}$$

$$\bar{c}: \; p_{c,t} \bar{c}_t = \omega y_t - s_t, t \in [2, T] \tag{69}$$

$$c_y: \; c_{y,t} = \phi p_{c,t} (\tilde{c}_t + \bar{c}_t), t \in [2, T] \tag{70}$$

$$q_k: \ \kappa_k \frac{k_t - k_{t-1}}{k_{t-1}} = q_{k,t} - 1, t \in [2, T] \tag{71}$$

$$q_d: \ \kappa_d \frac{d_t - d_{t-1}}{d_{t-1}} = q_{d,t} - 1, t \in [2, T] \tag{72}$$

$$r: \ r_t = (1-\pi)\left(\frac{x_{t+1}}{x_t}\right)^{\eta} - 1, t \in [2, T_Z] \tag{73}$$

Core:

$$\text{Uses: } k_t + \frac{\kappa_k}{2}\frac{(k_t - k_{t-1})^2}{k_{t-1}} + d_t + \frac{\kappa_d}{2}\frac{(d_t - d_{t-1})^2}{d_{t-1}} + c_{y,t} + \nu\gamma\frac{n_t}{q(n_t)} \tag{74}$$

$$\text{Sources: } y_t + (1-\delta_k)k_{t-1} + (1-\delta_d)d_{t-1} \tag{75}$$

$$\text{Uses} = \text{Sources }, t \in [2, T](T-1) \tag{76}$$

$$(1 + r_t)\mu_t = 1, t \in [2, T-1](T-2) \tag{77}$$

$$p_{k,t} = (1 + r_{t-1})(1 + f_{k,t})q_{k,t-1} - (1-\delta_k)q_{k,t}, t \in [3, T](T-2) \tag{78}$$

$$p_{d,t} = (1 + r_{t-1})(1 + f_{d,t})q_{d,t-1} - (1-\delta_d)q_{d,t}, t \in [3, T](T-2) \tag{79}$$

$$p_{d,t}d_{t-1} = (1-\phi)p_{c,t}(\tilde{c}_t + \bar{c}_t), t \in [2, T](T-1) \tag{80}$$

Total core equations: $2(T-1) + 3(T-2) = 5T - 8$

Unknowns: $T-1$ values of p_d and $\tilde{c}$ for $2T-2$, $T_Z - 1$ for n, $T-2$ values of k and d for $2T-4$, and $T - T_Z - 1$ for r, for a total of $5T-8$ unknowns.

Constant:

$$\bar{m} = \alpha\left(\frac{k^*}{\bar{n}}\right)^{1-\alpha} \tag{81}$$

C Sources Cited in Table 1

Hall (2004), Hall (2009b), Silva and Toledo (2008)