WHY HAS THE US ECONOMY RECOVERED SO CONSISTENTLY FROM EVERY RECESSION IN THE PAST 70 YEARS?

Robert E. Hall
Marianna Kudlyak

Working Paper 27234
http://www.nber.org/papers/w27234

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2020, Revised June 2021
Why Has the US Economy Recovered So Consistently from Every Recession in the Past 70 Years?
Robert E. Hall and Marianna Kudlyak
NBER Working Paper No. 27234
May 2020, Revised June 2021
JEL No. E32,J63,J64

ABSTRACT

A remarkable fact about the historical US business cycle is that, after unemployment reached its peak in a recession, and a recovery begins, the annual reduction in the unemployment rate is stable at around one tenth of the current level of unemployment. We document this fact in a companion paper, Hall and Kudlyak (2020a). Here, we consider explanations for the surprising consistency of recoveries. We show that the evolution of the labor market from recession to recovery involves more than the direct effect of persistent unemployment of job-losers from the recession shock -- unemployment during the recovery is above normal for people who did not lose jobs during the recession. We explore models of the labor market's self-recovery that imply gradual working off of unemployment following a recession shock. We emphasize the feedback from high unemployment to the forces driving job creation. These models also explain why the recovery of market-wide unemployment is so much slower than the rate at which individual unemployed workers find new jobs. The reasons include the fact that the path that individual job-losers follow back to stable employment often includes several brief interim jobs.

Robert E. Hall
Hoover Institution
Stanford University
Stanford, CA  94305-6010
and NBER
rehall@gmail.com

Marianna Kudlyak
101 Market St
Federal Reserve Bank of San Francisco
San Francisco, CA 94105
mkudlyak@gmail.com
Contents

1 Introduction 4

2 Uniform Unemployment Recovery across Recessions 7

3 Job Loss in Recessions 7
   3.1 Layoffs 8
   3.2 Mass layoffs 9
   3.3 Job destruction 11
   3.4 Displaced workers 12
   3.5 Comparison of measures of the spike of job loss in a recession 13
   3.6 Initial unemployment insurance claims 15
   3.7 Flow of new permanent layoffs in the Current Population Survey 17

4 The Direct Channel from Job Loss to Subsequent Lingering Unemployment 17
   4.1 Information about the subsequent role in unemployment from job displacement 17
   4.2 Application to other measures of job loss 20
   4.3 Conclusions about the relation between the magnitude of the increase in unemployment following a recession shock and the measures of job loss 20
   4.4 Excess unemployment of new entrants 21

5 Effective Exit Rate from Unemployment 22
   5.1 Defining and measuring the effective exit rate 22
   5.2 Implications of low effective job-finding rates 23

6 The DMP Model 26
   6.1 Potential driving forces of the DMP model 26
   6.2 Path of unemployment following a recession in the basic DMP model 28
   6.3 Path of unemployment in a model with low effective unemployment exit rate 31
   6.4 Variation of the driving forces over time 32

7 Models that Interpret Time Variation in the Driving Forces as Exogenous to the Labor Market 34
   7.1 Profitability of hiring a worker 34
   7.2 Financial sources of rising $P-W$ in recoveries 35
8 Endogenous Mechanisms Implying a Slow Downward Glide in Unemployment during Recoveries

8.1 Negative feedback from unemployment to tightness ........................................ 38
8.2 Vacancy costs ........................................................................................................ 40
8.3 Recruiting process and externalities ..................................................................... 41
8.4 Composition effects ............................................................................................... 42
8.5 Scarring effects ....................................................................................................... 43
8.6 The separation rate ................................................................................................. 43

9 Other Forces Operating during Recoveries

9.1 Fiscal and monetary policy ..................................................................................... 44
9.2 Other forces during recoveries ............................................................................... 46
9.3 Discussion of policies and other forces operating during recoveries .................. 49

10 Conclusions: The Sources of the Slow, Stable Downward Glide of Unemployment during Recoveries
1 Introduction

We study data from the labor market during recoveries from recessions, excluding the recovery from the pandemic recession of 2020. Our objective is to understand why the recovery phase of the US business cycle has invariably been slow but irresistible, a fact established in a companion paper, Hall and Kudlyak (2020a). The unemployment recovery process is similar in all of the past 10 recoveries: the annual reduction in the unemployment rate is stable at around ten percent of the prior level.

We note that a well-documented property of the unemployment rate is that unemployment rises rapidly in response to a significant aggregate adverse shock and then gradually recovers. Like fuel prices, unemployment rises like a rocket and falls like a feather. This property was most recently confirmed by Dupraz, Nakamura and Steinsson (2019), with many cites to the earlier literature.

Our principal claims are:

1. Recessions involve displacement of large numbers of workers, but the elevated level of unemployment along the recovery path involves far more people than the original displacement—unemployment is contagious.

2. Self-recovery occurs in the Diamond-Mortensen-Pissarides model even without any external force. But the recovery in the model with standard parameter values is much too fast, compared to data.

3. A model with negative feedback from unemployment to labor market tightness provides an internally consistent version of the DMP model with reliable but slow recoveries, as in the data. No external force is involved.

4. Sources of the negative feedback include cyclical changes in the composition of the unemployed, adjustment costs in vacancy creation, congestion in recruiting, scarring effects from lengthy unemployment, and persistence of elevated separation rate.

We focus on recoveries. Our measurement starts in an economy that has just been hit by an adverse shock that triggered a recession. This paper recognizes that the shocks that propel unemployment sharply upward are heterogeneous. The major recession that began in 1981 is generally viewed as the result of a sharp monetary contraction, while the major recession that began at the end of 2007 got much of its strength from the financial crisis of September 2008. Historical recoveries have been much more homogeneous.

We point out the puzzle of slow decline of unemployment. Cole and Rogerson (1999) first called attention to the puzzle—unemployment declines much more slowly than the measured exit rates from unemployment among individuals would seem to indicate.
We then ask, what accounts for the economy’s consistent, reliable record in recovering from adverse shocks? Our thesis is that the economy has a powerful tendency to self-recover from serious adverse shocks, but recovery takes time.

We consider negative feedback from high unemployment to the job finding rate as a key mechanism behind the slow unemployment recoveries. Our discussion of unemployment is within the framework of Diamond, Mortensen, and Pissarides (DMP). Their model has a well-known but counter-intuitive property—it lacks feedback from unemployment to labor-market tightness. When an adverse shock creates a high volume of unemployment, but the shock subsides so the determinants of tightness return to normal, the legacy of unemployment has no discouraging effect on tightness. Jobs are just as easy to find with unemployment at 10 percent as they are when unemployment is 4 percent. Much of this paper is devoted to studying modifications of the DMP model to alter this property. In the modified model, unemployment is much more persistent, because jobs are hard to find when unemployment is high.

Our view of the labor market has points in common with Pries (2004) and can be seen as responding to the challenge of Cole and Rogerson (1999) to explain why aggregate unemployment recovers much more slowly than does an individual spell of unemployment.

We proceed in the following steps: First, we study the job loss that occurs when a crisis launches a recession. A spike in job loss is visible in a variety of data sources that measure layoffs, job destruction, displacement, and unemployment insurance claims. But the spike in job loss is short-lived as compared to ensuing elevated unemployment.

Second, we ask whether the volume of job losers and their likely speed of finding long-term replacement jobs is enough to explain the long-lasting bulge of total unemployment that is only gradually worked off during even a long recovery like the one that ended in early 2020. We conclude, from data on displaced workers collected every two years in the Current Population Survey, that the number of workers displaced even in the severe recession starting in 2007 was not enough to explain the volume of excess unemployment present in the US economy during the period from 2009 through 2014. Something happened in the labor market during that period that caused elevated unemployment among workers who were not displaced around 2009. Unemployment proved to be infectious.

Third, we examine the puzzle of low recovery speed in the framework of the DMP model. We calculate the effective exit rate from unemployment, which is lower than the exit rate for individuals from one month to the next. Those individual exits are frequently temporary departures from the labor force or short-term jobs, and are then followed by additional spells of unemployment, as described in Hall and Kudlyak (2019). Short-term jobs are an important part of the job-finding process, as studied earlier in Hall (1995). In the DMP
equilibrium with our estimated low effective exit rate, the unemployment rate falls more gradually, so it accounts for some of the puzzle of low recovery speed.

We study models that explain slow but sure recoveries through feedback from the level of unemployment to the job-finding rate. We consider the feedback to various driving forces. For example, when unemployment spikes, employers’ costs of recruiting rise. According to standard DMP principles, higher costs of filling vacancies discourage job creation and raises equilibrium unemployment. This model generates a generally slow decline of unemployment during a recovery. We consider mechanisms based on endogenous feedback from unemployment to vacancy creation costs. We review the extensive literature on the feedback from unemployment to the recovery process that rebuilds employment lost in the earlier crisis. Another idea regarding slowing down the unemployment recovery in the DMP model is Fujita and Ramey’s (2007) model with costs of adjustment of vacancies.

We describe a wide range of other mechanisms that participate in the gradual reduction of unemployment during a recovery. These include

1. A gradual return to the normal mix of unemployment, away from the disproportionate role of hard-to-re-employ workers in the aftermath of a recession
2. Slow but reliable decline in labor-market churn that occurs following a recession
3. Decline back to normal from strict credit standards put in place during a recession
4. Congestion effects impeding recruiting efforts when unemployment is high

We conclude that the economy includes a strong internal force toward recovery that operates apart from policy instruments and apart from productivity growth and financial developments revealed in the stock market. After a negative shock, employers gradually find it profitable to hire more aggressively. Unemployment falls as the unemployed are put back to work. Rather than a pull from expansionary policy, the growth in employment arises from a push toward lower unemployment.

We find that only a small part of the reduction in unemployment following a recession takes the form of the first jobs found by the workers who lost jobs as a direct result of the recession. Additional spells of joblessness occur within that group, and extra unemployment occurs among workers who were not immediate victims, such as people entering the labor market for the first time, after the recession. This induced unemployment gradually returns to normal in the recovery.

This paper is about recoveries from January 1948 through February 2020 and not about the recovery from the pandemic that influenced the labor market starting in late March 2020. But the principles studied in this paper apply to the recovery from the pandemic recession as
well if one properly accounts for the burst in unemployment due to temporary layoffs. That pandemic created an unprecedented increase in the non-working population—unemployed and out of the labor force. But many of these people had good prospects of recall to their earlier jobs or successful re-entry to the labor force once the pandemic ended. In the previous recessions, the fraction of the unemployed on temporary layoff was small (Hall and Kudlyak (2020b)). The unemployed on temporary layoff typically do not need to go through the costly and time-consuming process at the center of the mechanism studied in this paper. The principles studied in this paper apply to the rest of the unemployed.

Why has the US economy recovered so consistently from every recession in the past 70 years? Our answer: Recoveries are endogenous—there is a natural force causing job-seekers to match with available jobs and to lower unemployment. The bulge of unemployment created by a crisis at the beginning of a recovery creates a negative feedback to labor market tightness, endogenously slowing the recovery.

Our results have implications for macroeconomic policy: During a recovery, unemployment seems little responsive to demand disturbances that do not throw the economy into recession. Economic policy should focus on preventing recessions rather than trying to ameliorate their effects as recoveries proceed slowly but reliably.

2 Uniform Unemployment Recovery across Recessions

In Hall and Kudlyak (2020a), we study US business-cycle recoveries over the past 70 years. We focus on the unemployment rate. Our key results are (1) the recovery process takes place reliably, regardless of the nature of the shock that causes the preceding economic contraction, and (2) the recovery process is similar in all of the ten past recoveries—unemployment falls by about 0.1 log points per year.

Figure 1 displays the log of the unemployment rate during the 10 recoveries since 1948, with the recession spells of sharply rising unemployment left blank. Throughout the paper we exclude the recovery from the pandemic recession that started in 2020.

The key fact about recoveries is apparent in the figure: Unemployment declines smoothly but slowly throughout most recoveries most of the time, at close to the same proportional rate. In the log plot, the recoveries appear as impressively close to straight lines.

3 Job Loss in Recessions

In this section, we examine job losses in recessions. Data on layoffs, job destruction, and long-term worker displacement show the substantial but short-lived spikes of job loss when
the aggregate economy contracts. The worker-level data from the Current Population Survey and data on the initial unemployment insurance claims show the initial substantial spike and a subsequent lingering of the elevated job loss.

We consider a number of measures of job loss:

- Layoffs, the flow of workers whose jobs ended at the initiative of employers.
- Job destruction, the amount of employment decline among establishments with shrinking employment.
- Worker displacement, job loss among workers with at least three years of tenure at the lost job.
- Unemployment insurance claims.

### 3.1 Layoffs

Figure 2 shows data on layoffs from the Job Openings and Labor Turnover Survey. A layoff occurs when an employer terminates a worker without prejudice, typically because continuing employment has become unprofitable. Most layoffs occur without any definite promise to rehire, but explicitly temporary layoffs are an important part of layoffs. On average, 20
million workers lose their jobs each year in normal times. A substantial but short-lived burst of above-normal layoffs occurred soon after the financial crisis in the fall of 2008.

### 3.2 Mass layoffs

A mass layoff occurs when a relatively large number of a firm’s employees lose their jobs. These events include high-tenured workers who tend to suffer prolonged periods of joblessness following job loss (Jacobson, LaLonde and Sullivan (1993), Davis and von Wachter (2011)).

The Mass Layoffs Statistics program of the Bureau of Labor Statistics tracks the effects of major job cutbacks using data from state unemployment insurance databases. A mass layoff is defined as 50 or more initial claims for unemployment insurance benefits being filed against an employer during a 5-week period. These employers are contacted by the state agency to determine whether the separations lasted more than 30 days. Such events are termed extended mass layoffs. The BLS obtains information on the total number of workers separated during the extended mass layoffs, including the workers who do not file for unemployment insurance, and the reasons for these separations according to the employer. These layoffs involve both people subject to recall and those who are terminated. The program operated from 1995 to the first quarter of 2013.

Figure 3 shows the number of initial claimants from extended mass layoffs. The number hovers around a million in normal times but spikes during recessions. A decline in business
demand and financial difficulties are the main reasons cited behind the spikes. In 2009, extended mass layoffs spiked to 2.4 million.

Another source of data on mass layoffs is the Worker Adjustment and Retraining Notification Act (WARN), which requires employers to provide notice 60 days in advance of covered plant closings, covered mass layoffs, or sale of business that result in an employment loss. Employers are covered by WARN if they have 100 or more employees, not counting employees who have worked less than 6 months in the last 12 months and not counting employees who work an average of less than 20 hours a week. The term employment loss means (1) an employment termination, other than a discharge for cause, voluntary departure, or retirement; (2) a layoff exceeding 6 months; or (3) a reduction in an employee’s hours of work of more than 50 percent in each month of any 6-month period. A plant closing occurs if an employment site will be shut down, and the shutdown will result in an employment loss for 50 or more employees during any 30-day period. A mass layoff occurs without a plant closing if the layoff results in an employment loss at the employment site during any 30-day period for 500 or more employees, or for 50-499 employees if they make up at least 33 percent of the employer’s active workforce. Under certain circumstances, smaller employment losses also trigger notification requirements.
Figure 4: Mass Layoffs, by State

Note: Data from the layoff notices under the Worker Adjustment and Retraining Notification Act.

The WARN data over an extended period of time are publicly available for many states. Figure 4 shows the number of layoffs for Alabama, Michigan, and Washington, as examples. The data show clear spikes in layoffs in 2009. For Alabama and Washington the figure shows layoffs sorted by the effective date. For Michigan, we have information about the date of the WARN notice but not about the effective date of the layoff, so we sort the layoffs by the expected effective date, which is the notice date plus two months.

3.3 Job destruction

The Business Dynamics Statistics data report job destruction. This measure is defined as the sum of all establishment-level reductions in employment. Davis and Haltiwanger (1992) and Davis, Faberman and Haltiwanger (2013) proposed job destruction as a measure
of separations and validated the definition through study of the microdata from JOLTS. Although an employer could accomplish a reduction in employment by cutting back hiring and relying on normal attrition, in fact, almost all employment reductions take the form of separations. When an adverse shock hits the economy, separations jump and quits fall.

Figure 5 shows data from the BDS on job destruction. It shows a considerable bulge of job destruction immediately after the financial crisis.

3.4 Displaced workers

Displaced workers are defined as those 20 years old and over who have worked for their employers for 3 or more years at the time of displacement, who lost or left jobs because their plants or companies closed or moved, because there was insufficient work for them to do, or because their positions or shifts were abolished. These are job losses among workers with substantial tenure, in contrast to layoffs measured in JOLTS. These are called long-tenured displaced workers.

Table 1 shows the findings of the displaced workers supplement to the CPS taken in January of even-numbered years starting in 2002. The survey inquires about current un-
<table>
<thead>
<tr>
<th>Survey in January of Displacement occurring in calendar years</th>
<th>Number of displaced workers</th>
<th>Unemployed at time of survey</th>
</tr>
</thead>
<tbody>
<tr>
<td>2002 1999 2000 2001</td>
<td>3,969</td>
<td>841</td>
</tr>
<tr>
<td>2004 2001 2002 2003</td>
<td>5,329</td>
<td>1,076</td>
</tr>
<tr>
<td>2006 2003 2004 2005</td>
<td>3,815</td>
<td>511</td>
</tr>
<tr>
<td>2008 2005 2006 2007</td>
<td>3,641</td>
<td>655</td>
</tr>
<tr>
<td>2010 2007 2008 2009</td>
<td>6,938</td>
<td>2,505</td>
</tr>
<tr>
<td>2012 2009 2010 2011</td>
<td>6,121</td>
<td>1,634</td>
</tr>
<tr>
<td>2014 2011 2012 2013</td>
<td>4,292</td>
<td>893</td>
</tr>
<tr>
<td>2016 2013 2014 2015</td>
<td>3,191</td>
<td>507</td>
</tr>
<tr>
<td>2018 2015 2016 2017</td>
<td>2,981</td>
<td>429</td>
</tr>
</tbody>
</table>

Table 1: Total Long-Tenured Displaced Workers and the Number of Unemployed at the Time of the Survey, in Thousands

Note: Data from the Worker Displacement Supplement to the Current Population Survey.

...employment and displacement in the year ended the month before, and one and two years earlier.

The design of the displaced workers supplement to the CPS poses an interesting challenge to inference about the time path of long-term displacements and the path of unemployment following displacement. Figure 6 shows an attempt. The annual estimates satisfy the overlapping three-year sums and are informed by the timing of layoffs and job destruction within each three-year span. The figure also shows a counterfactual path of displacements, which eliminates the two recession spikes present in the actual data.

### 3.5 Comparison of measures of the spike of job loss in a recession

Figure 7 compares the estimated long-term worker displacement counts to the tabulations of layoffs, job destruction, and extended mass layoffs. Although the normal level of displacement is far below the levels of layoffs or job destruction, the increase in displacements at the outset of the two recessions is an important fraction of the increases for layoffs and job destruction.

Figure 8 shows excess job loss associated with the 2009 recession by four measures of job loss, together with excess unemployment. We measure excess job loss as the job loss in excess of the average job loss just before and just after the year of the job loss spike.
Figure 6: Estimated Annual Displacements, Actual and Counterfactual

Figure 7: Annual Levels of Layoffs, Job Destruction, Displacements, and Extended Mass Layoffs, in Thousands
We measure excess unemployment, as unemployment in excess unemployment in 2007. All four job loss measures show a substantial but short-lived spike. Unemployment shows a substantial increase and slow return to its pre-recession level.

3.6 Initial unemployment insurance claims

Figure 9 shows initial unemployment insurance (UI) claims. In contrast to layoffs (Figure 2) but similarly to unemployment, during recessions, the initial UI claims go up like a rock and down like a feather.

Why is there a discrepancy between the number of layoffs and the number of initial UI claims? One factor is that not all eligible unemployed individuals claim the benefits. Building on Blank and Card (1991), Auray, Fuller and Lkhagvasuren (2019) find that from 1989 through 2012, the take-up rate averaged 77 percent. Research shows that the number of those who are eligible but do not claim benefits increases in recession and declines in recoveries (see Fuller, Ravikumar and Zhang (2012) and Auray et al.). Thus, fluctuations in take-up rates goes in the wrong direction as an explanation of the discrepancy between layoffs and the initial UI claims.
Figure 9: Initial Unemployment Insurance Claims, in Thousands

Figure 10: Unemployment due to Permanent Job Loss, by Duration

Note: Data from the Current Population Survey.
3.7 Flow of new permanent layoffs in the Current Population Survey

Figure 10 shows unemployment involving permanent job loss, by duration, from the CPS. Layoffs with duration of 5 weeks or less is the flow of new layoffs. The flow spikes at the onset of recessions and declines only slowly afterwards.

4 The Direct Channel from Job Loss to Subsequent Lingering Unemployment

We consider the hypothesis that excess job loss directly accounts for the spike and the subsequent long slow decline of excess unemployment. We call it the direct-channel hypothesis. According to this hypothesis, the extra individuals who become unemployed because of the recession shock follow a path similar to those found in research such as Jacobson et al. (1993) and Davis and von Wachter (2011) that tracks the post-displacement paths of workers who lose their jobs from layoffs. These paths often include multiple spells of unemployment.

4.1 Information about the subsequent role in unemployment from job displacement

The CPS survey supplement measuring job displacement contains crucial information about lingering unemployment among job-losers in the years following job loss.

We fit a simple time-series regression with the biennial data for unemployment in January of even-numbered years of workers suffering displacements in the previous three years as the left-hand variable and three lagged values of the estimated displacement counts as right-hand variables, along with a constant. The relation takes the form

\[ u_t = f_1(D_{t-1}) + f_2(D_{t-2}) + f_3(D_{t-3}) \] (1)

We linearize as

\[ u_t \doteq \alpha + \beta_1 D_{t-1} + \beta_2 D_{t-2} + \beta_3 D_{t-3} \] (2)

If market tightness were constant over time, \( \beta_1 \) would be the unemployment rate among workers who suffered displacement within the past year, \( \beta_2 \) one to two years ago, and \( \beta_3 \) two to three years ago. The design of the survey prevents learning about unemployment among people displaced more than 3 years ago. However, job-finding rates are lower in the same years that displacements are high, so \( f(D) \) is a convex function of \( D \). This property implies that the intercept \( \alpha \) should be negative and the coefficients should be greater than the unemployment rates.
### Table 2: Regression Results for the Relation between Lagged Displacements and Current Unemployment of Workers Suffering those Displacements

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Interpretation</th>
<th>Coefficient</th>
<th>Standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>Intercept</td>
<td>-991</td>
<td>(144)</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>Effect of prior year's displacements</td>
<td>0.76</td>
<td>(0.07)</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>Effect of displacement 1 to 2 years ago</td>
<td>0.21</td>
<td>(0.08)</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>Effect of displacement 2 to 3 years ago</td>
<td>0.37</td>
<td>(0.06)</td>
</tr>
</tbody>
</table>

$R^2 = 0.991$

$\sigma = 116$

Table 2 shows the regression results. The good fit suggests that the imputation of annual timing for the displacements is reasonably successful. The fact that the 3rd-year coefficient is somewhat larger than the 2nd-year one is within sampling variation, but may also reflect the fact that a worker with displacement 3 years ago also suffered an earlier displacement as well. In addition, there may be a stronger convexity effect for the 3rd-year displacement. The negative intercept confirms the expectation of a convex relation between displacements and later unemployment.

Our first use of the regression is to impute unemployment of workers suffering displacements in the previous three years in January of the odd-numbered years when the supplement to the CPS was not performed. Figure 11 shows the fitted values from the regression for the years 2002 through 2018, in red, along with the actual unemployment counts for the even-numbered years when the supplement to the CPS occurs, in blue.

Our second use of the regression results is to calculate how much lower displacement-related unemployment would have been absent the spikes of displacement in the two recessions. Figure 12 shows the results of this counterfactual and compares displacement-related unemployment from the two recessions to overall unemployment in January in the years since 2001. The rise after the recession is material relative to the overall increase in unemployment following the recession that began in 2001, but is a small part of the large increase in unemployment following the financial crisis.
Figure 11: Actual and Fitted Unemployment Counts of Workers Suffering Displacements in the Previous Three Years, in Thousands

Figure 12: Total Unemployment and the Component Related to Bursts of Displacement Associated with Two Recessions, in Thousands
4.2 Application to other measures of job loss

We use the estimates from Table 2 to calculate excess unemployment from excess job loss by the four measures shown in Figure 8. Figure 13 shows the unemployment resulting from excess job loss in 2009 and total excess unemployment.

Figure 14 shows the contribution of unemployment from excess job loss in 2009 to the cumulative excess unemployment during the 2007-09 recession.

4.3 Conclusions about the relation between the magnitude of the increase in unemployment following a recession shock and the measures of job loss

An unambiguous spike in regular and mass layoffs, job destruction, and displacement, and mass layoffs accompanies the shock that marks a recession. Figure 13 shows that the spike in layoffs more than fully accounts for the spike in unemployment in 2009 but cannot account for all of the excess unemployment afterwards. That is, the excess job loss accounts for the magnitude of the initial increase in unemployment, but not its persistence. The persistence is too large to be explained as reflecting only the personal experiences of the extra joblosers dating from the spike. The direct channel is only part of the story of persistent high unemployment after the crisis. This conclusion is reinforced by the rise in unemployment among new entrants to the labor force, as we show below.
Figure 14: Contribution of Excess Job Loss to Cumulative Excess Unemployment, by Four Measures of Job Loss, in Thousands of Worker-Years

The results show that the spike of job loss during recessions induces a downstream effect on unemployment. Some of subsequent unemployment comes from the unemployed who suffered the original displacement event and are circling through short employment spells, and the rest of subsequent elevated unemployment appears to come from new job loss, not associated with the original job loss.

4.4 Excess unemployment of new entrants

By definition, new entrants to the labor force are not the victims of job loss events. A bulge of unemployment of new entrants following an adverse shock indicates that unemployment is infectious or the bulge of unemployment arises from a decline in the incentives to create jobs. Figure 15 shows that new-entrant unemployment nearly doubled after the financial crisis. This finding rules out the hypothesis that the sole cause of lingering unemployment following the crisis was the slow absorption of workers who suffered job loss from the crisis. The direct channel cannot be the only link between a crisis and its subsequent gradual recovery.
5 Effective Exit Rate from Unemployment

5.1 Defining and measuring the effective exit rate

From the data on unemployment in the displaced workers supplement, we can estimate what we call the effective exit rate from unemployment, denoted \( f_t \). We know the number of people in the survey who were displaced in the prior three years and who are currently unemployed. We also have our estimates of the number of people displaced in each of those years. The effective exit rate is based on the assumption that the probability of being unemployed \( \ell \) months following a displacement is the product of the monthly exit rates from the time of displacement up to the survey. Here we adopt the perspective originated in Krueger, Cramer and Cho (2014), and expanded in Hall and Schulhofer-Wohl (2018) and Hall and Kudlyak (2019), that the typical path from initial unemployment to current labor-market activity often involves a mixture of spells of short jobs, time out of the labor force, and unemployment. As Krueger and co-authors showed, the probability of being unemployed a year later conditional on starting unemployed is much higher than would be expected from the monthly probability of unemployment ending raised to the 12th power. Our calculation
here extends the calculation by two additional years, as we exploit the 3-year look-back in the displaced workers supplement of the CPS.

The implied relation between the observed number of people unemployed in the January survey of month $t$ is

$$
U_t = \sum_{\ell} \prod_i (1 - f_{i-\ell}) N_{\ell}.
$$

We parametrize as

$$
f_{\tau} = a - bu_{\tau}.
$$

The parameter $b$ is the negative sensitivity of the effective exit rate to the standard national unemployment rate $u$. Not surprisingly, it turn out to be essentially 1. We estimate $a$ and $b$ by minimizing the sum of squared residuals of the actual values to the implied values of $U_t$. The estimated value is $b = 1.00$ and the monthly effective exit rates range from 0.042 in 2010 to 0.099 in 2018. By contrast, the monthly exit rate is around 0.5.

### 5.2 Implications of low effective job-finding rates

This subsection provides evidence that recessions are followed by long periods of high but not continuous unemployment among those who lost jobs in the recession. During the long re-employment process, the unemployed often circle among unemployment, out of the labor force, and short-term jobs.

Figure 16 shows unemployment by reason as a share of the labor force, except for labor force new entrants. The figure shows that recessions involve not only an increase in unemployment from permanent and temporary layoffs but also due to completion of temporary jobs and labor force re-entry. This points towards an elevated number of individuals taking temporary jobs and circling between unemployment and spells out of the labor force.

When a crisis causes a spike in unemployment, there is a shift away from stable jobs and toward brief jobs in the working-age population. This shift gradually subsides during the recovery. To demonstrate this phenomenon, we study the 8-month CPS activity paths, as in Hall and Kudlyak (2019). We define short employment spells as those lasting one or two months. These are the spells that are preceded and succeeded by unemployment or out of the labor force. We define an individual to have stable employment if employed in all 8 reported months. We calculate the average number of short employment spells among the of CPS respondents of working age. We also calculate the average number of respondents in stable employment. We create an index of the shift toward short jobs as the difference between first and second of these calculations. Figure 17 shows the short-spell index starting in 1976 for four demographic groups. The indexes jump upward in recessions and gradually declines during the ensuing recovery for all four groups.
Figure 16: Unemployment by Reason, as Share of Labor Force

Note: Data from the Current Population Survey.
Figure 17: Indexes of the Duration of Employment Spells

Note: Authors calculations using data from the Current Population Survey.
6 The DMP Model

This section describes the basic DMP model and discusses potential mechanisms and driving forces in the model that may help explain the slow but steady recovery of unemployment following a spike in job loss. Our goal is to explain some key features of the mechanism of the DMP model, not to create a new model at the state of the art. Put another way, we discuss the model of Chapter 1 of Pissarides (2000) and its key mechanism, which is at heart of the models of later chapters.

The DMP class of models treats the level of unemployment as a state variable. At the outset, in our application of the model, unemployment has a high value left behind by a recession. The model traces the movements of unemployment for a decade, aiming to emulate the slow but steady decline of unemployment documented in Hall and Kudlyak (2020a). We discuss potential driving forces of the model that generate persistence. We confirm the well-known failure of the simple DMP model to account for the slow pace of recovery without invoking persistent movements of driving forces.

6.1 Potential driving forces of the DMP model

Our discussion of the model will make the simplifying assumption that the only flow in the labor market that is sensitive to tightness is from unemployment to employment. To keep the exposition compact for now, we take the size of the labor force to be a constant. Thus we neglect variations in flows into and out of the labor market. Another important simplification is that we omit on-the-job search. A large and rich literature deals with these topics (see, for example, Krusell, Mukoyama, Rogerson and Sahin (2017). As we discussed in the section above, movements in and out of the labor force are an important and integral part of the job-finding process after a job loss in recessions. However, to explain the key insights of our paper, it suffices to focus on a simple textbook model with employment and unemployment. Our calibration of the model to the low effective exit rate from unemployment serves as a reduced-form representation of the churn that an unemployed worker goes through in order to find a long-term job.

The labor market operates on the principles of random search. The tightness of the market, denoted \( \theta \), is measured by the ratio of vacancies to job-seekers. We normalize the labor force at 1 and measure vacancies as the ratio to the labor force, denoted \( v \), and unemployment as the unemployment rate, denoted \( u \), so

\[
\theta = \frac{v}{u}. \tag{5}
\]
An aggregate constant-returns matching function $m(u, v)$ gives the flow of job matches as a function of the inputs—the stock of searchers, $u$, and the stock of vacancies $v$. We parametrize $m(u, v)$ as a Cobb-Douglas function $m(u, v) = \mu u^{1/2}v^{1/2}$. The parameter $\mu$ is matching efficiency, analogous to Hicks-neutral technical change.

The job-finding rate is the number of matches per job-seeker per month,

$$f = \mu \frac{u^{1/2}v^{1/2}}{u} = \mu \theta^{1/2}. \quad (6)$$

Thus, tightness determines the monthly job-finding rate, an increasing function. Similarly tightness determines the monthly job-filling rate, $q = \mu \theta^{-1/2}$. The latter is the number of jobs filled by holding a vacancy open for a month, so it can be greater than 1. The job-filling rate decreases with tightness.

We let $P$ be the present value of a newly hired worker’s productivity and $W$ be the present value of their wage. The difference $P - W$ is the net benefit to the firm from hiring a new worker. For simplicity, we call $W$ the wage but we mean the present value the worker earns from the job, as of the time of hiring. The model operates in an environment of certainty, so expectation operators are omitted.

The flow cost of recruiting is $\kappa$. Recruiting satisfies the zero-profit condition,

$$\kappa = \mu \theta^{-1/2}(P - W) \quad (7)$$

This condition pins down tightness:

$$\theta = \left(\frac{P - W}{\mu \kappa} \right)^2 \quad (8)$$

Tightness is increasing in matching efficiency $\mu$, increasing in productivity $P$, decreasing in the wage, $W$, and decreasing in the recruiting cost, $\kappa$.

The law of motion of unemployment describes the rate of change of unemployment as the net of inflows from separations and outflows from job-finding:

$$\dot{u} = (1 - u)s - u \mu \theta^{1/2}. \quad (9)$$

The parameter $s$ is the separation rate into unemployment. Its reciprocal, $1/s$, is the expected duration of a job. In addition to its visible role in the law of motion, the separation rate is one of the determinants of the present values $P$ and $W$.

Separations matter in two ways. First, a shock that hits the economy just before the starting time of the model results in a pulse of separations and an elevated unemployment rate at the outset of the time span of the model. We do not model the shocks or recessions—they are simply the source of a legacy of unemployment when the model swings into action.
Second, the separation rate $s$ describes the flow of separations during the recovery. The separation rate controls the inflow to unemployment and the job-finding rate controls the outflow. Along a realistic recovery path, the two flows are almost equal—unemployment declines quite slowly. For now, we take the separation rate to be an exogenous constant.

The job-finding rate, $\mu \theta^{1/2}$, controls the flow out of unemployment. Its reciprocal is the expected duration of a spell of unemployment. Note that matching efficiency $\mu$ appears in both the tightness equation and in the law of motion for unemployment.

The potential driving forces of the model are

- present value of a newly hired worker’s productivity, $P$,
- present value of their wage, $W$,
- flow cost of a vacancy, $\kappa$,
- matching efficiency, $\mu$,
- separation rate, $s$.

With these specified as constants, time series, or functions, the model is a first-order differential equation in the single state variable, $u$. Here and in the rest of the paper, we refer to driving forces, which are variables taken as exogenous to the labor market, in the sense that we do not consider that actors in the labor market can influence the variables.

We consider the path of the economy immediately after a shock has left unemployment at an elevated level. The model then evolves according to its law of motion. In the cases we consider, it converges to a stationary state because the driving forces approach constant levels.

### 6.2 Path of unemployment following a recession in the basic DMP model

We start by describing a basic DMP model, with constant productivity, separation rate, matching efficiency and vacancy cost. We also assume that the wage is constant even though the unemployment rate declines over time. This assumption mirrors the behavior of the canonical DMP model of Mortensen and Pissarides (1994). In that model, the wage is the endogenous result of bargaining. Under our assumptions of constant $P$ applied to that model, the bargained wage is constant, even though unemployment follows a path that starts above its stationary value and converges over time to its stationary value. Because the wage is constant, all of the driving forces are constant and labor-market tightness is constant during the recovery.
In the data, tightness rises during recoveries. Modifying the DMP model to match this key fact is one of the main topics of recent theoretical work on the model and of this paper. We parametrize the model to resemble the economy in early 2020. Time is monthly. The stationary unemployment rate is $u^* = 0.035$, separation rate $s = 0.018$, and matching efficiency $\mu = 0.5$.

Figure 18 describes the behavior of the model in a phase diagram, with unemployment $u$ on the horizontal axis and tightness $\theta$ on the vertical axis (see Pissarides (2000), Figure 1.3, p. 30). Equation (8) determines tightness. It describes hypothetical combinations of tightness and unemployment. It is a horizontal line in the phase diagram because in the basic model tightness is a constant, not a function of unemployment.

The downward-sloping curve in the phase diagram is the locus of stationary values of unemployment, derived from the law of motion by setting $\dot{u}$ to zero. It is

$$\theta = \left(1 - \frac{u s}{u + \mu}\right)^2. \quad (10)$$

We call this the “$\dot{u} = 0$ curve” throughout the paper. It appears in all of our phase diagrams. It only changes when we consider different values of matching efficiency $\mu$ or the separation rate $s$. Points above and to the right of the $\dot{u} = 0$ curve have declining unemployment, while points down and to the left have rising unemployment. At high unemployment rates, a given job-finding rate generates a higher outflow from unemployment because the rate applies to more people. With lower unemployment, constancy of unemployment along the locus requires a higher job-finding rate and thus higher tightness. Or, put another way, higher tightness means a higher rate of growth of unemployment at a given unemployment rate, and so a lower level of unemployment to achieve constancy.

All the combinations of unemployment and tightness consistent with the model will be on the horizontal line labeled $\theta$, tightness. After a recession creates a legacy of high unemployment, the economy starts at the right end of that line. As the recovery proceeds, unemployment moves horizontally to the left according to equation (9).

Figure 19 shows the path of unemployment implied by the simple model with constant driving forces, along with the actual path starting from the peak in 2009. The dots show the progress by month. In the first month, unemployment falls substantially. As the unemployment rate falls during the recovery, the steps become smaller. The model economy closes most of the gap in just three or four months. The model’s recovery is far speedier than actuality, the point made emphatically by Cole and Rogerson (1999).

All of our model solutions in this paper are effectively exact, that is, not based on any approximation, and using double-precision arithmetic. Petrosky-Nadeau (2014) demonstrated
Figure 18: Phase Diagram for the Basic DMP Case
The labor market begins at the right end of the horizontal line, with high unemployment. The market moves to the left, along that line, as unemployment falls but tightness remains the same. The market approaches the left end of the line where unemployment becomes constant at its stationary level.

Figure 19: Recovery Path of Unemployment in the Standard DMP Model with Constant Driving Forces, and Actual Unemployment
the importance of accurate solutions of DMP models, arising from the substantial concavity of the matching function, which impairs the accuracy of approximation by log-linearization.

6.3 Path of unemployment in a model with low effective unemployment exit rate

We know that the 50-percent per month transition rate from unemployment to employment in the basic DMP model greatly overstates the actual exit rate from unemployment (see section 5). Figure 20 shows the phase diagram and Figure 21 shows the model’s unemployment path together with the actual path, with the lower effective exit rate from unemployment of 0.1 per month, in the range estimated in Section 5. This alteration substantially delays the recovery but not nearly enough to match the actual path of unemployment. The phase diagram with lower exit rate is same as the earlier one, except that the dots are closer to each other, so we do not repeat it.

We conclude that using the estimated effective unemployment exit rate of 0.10 makes an important contribution to matching the model’s unemployment path but cannot be a full resolution of the high-persistence puzzle. Consequently, short-term jobs (see, for example, Hall (1995)) and movements between unemployment and out of the labor force, which are captured by the low effective exit rate, are an important ingredient of slow recoveries of unemployment but cannot fully explain it.
Pries (2004) builds a DMP model that explains the high persistence of unemployment as the result of recurrent spells of unemployment following a shock. Once a job match is made, the parties are at risk of an adverse productivity realization that reveals that the match should end and the worker should return to the labor market. In normal times, most matches will have become known to be reliable and no longer at risk of being found unproductive. At random, a cloud may form over the labor market that calls into question the earlier belief that a match is good—the parties need to receive a new signal of reliability for a fraction of the existing matches. Some of the matches end immediately and the others are exposed to the possibility that they will be found to be unproductive from a later draw of productivity. If the aggregate shock simply knocks out some of the existing matches, the model generates little persistence—the victims of the shock regain reliable employment almost as quickly as they would without the learning-by-experience feature of the model (see Pries’s Figure 3). The broader version of the shock, which induces the parties to wait to determine who are the job losers, makes the effect of the shock realistically persistent.

6.4 Variation of the driving forces over time

So far we have discussed the model’s dynamics when the driving forces are constant. The basic model generates speedy recovery of unemployment. Tightness is fixed.
In this subsection, we consider evidence of time variation in the driving forces. Where possible, we use evidence from the direct empirical counterparts of the driving forces. Otherwise, we infer variation in a combination of driving forces using the model.

Labor-market tightness, $\theta$, has five driving forces—matching efficiency, $\mu$, productivity, $P$, the wage, $W$, the flow cost of a vacancy, $\kappa q$, and the separation rate, $s$. We infer variation in matching efficiency from the matching function and data on hires, vacancies, and unemployment. It is the ratio of the hiring flow to weighted matching inputs:

$$\mu = \frac{H}{U^{1/2}V^{1/2}}.$$  \hfill (11)

The other determinants of $\theta$—productivity, wage, and the vacancy cost—are not easily separately identified. We can identify the compound force, the job value, $J = (P - W)/\kappa$. From the zero-profit condition, this turns out to be the ratio of vacancies to hires:

$$J = \frac{\theta^{1/2}}{\mu} = \frac{\sqrt{V/U}}{H/\sqrt{UV}} = \frac{V}{H}.$$  \hfill (12)

Figure 22 shows the results. The labor market tightened during the long expansion starting in 2009 from a combination of rising matching efficiency $\mu$, and rising value of the compound job-creation incentive $J$.

The remaining driving force is the separation rate into unemployment, $s$. Figure 23 shows the time variation in the separation rate constructed from CPS data. The separation

Figure 22: Matching Efficiency, $\mu$, and the Compound Driving Force, $J$, Embodying Productivity, Wage, and Vacancy Cost
rate increases sharply in recessions and declines slowly in recoveries. The separation rate contributes negatively to tightness, so its decline in recoveries contributes to the gradual decline in unemployment.

7 Models that Interpret Time Variation in the Driving Forces as Exogenous to the Labor Market

Next we explore time-varying driving forces of the DMP model that are taken as exogenous to the labor market. They would be endogenous in a full general-equilibrium model.

7.1 Profitability of hiring a worker

The financial incentive to engage in job creation, $J = (P - W)/\kappa$, is at the center of the DMP class of unemployment models. If $J$ remains at a low value after a crisis and only gradually trends upward during the recovery, tightness will gradually rise from a low value.
back to normal, and unemployment will gradually decline. A large fraction of the DMP literature invokes this mechanism to explain movements of tightness.

This approach typically takes $P$ as an exogenous time-series process and $W$ as an endogenous variable. Unless $W$ changes by the same amount as $P$, an unlikely configuration, $P - W$ inherits some of the movements in $P$ and so tightness will rise in a recovery. But, in fact, productivity has essentially no correlation with tightness, so this channel cannot explain the cyclical movements of tightness and unemployment, with persistent recoveries (Hall (2017)).

A straightforward way to increase volatility in tightness relative to the volatility in $P$ is through a rigid wage, $W$. The wage is the result of a bargaining process and is generally endogenous, although a fixed wage is an interesting special case. Wage-bargaining models different from the Nash bargain of the canonical DMP model can deliver realistic volatility but do not generally add to persistence of recoveries. Shimer (2005) found that the movements in $P - W$ induced by movements in $P$ and thus in unemployment, were tiny, in the DMP model with Nash bargaining under plausible assumptions about parameter values. Numerous subsequent papers altered the original model to boost its response to productivity. A different and more fundamental shortcoming of the hypothesis that productivity drives unemployment is the lack of any correlation of measured productivity and unemployment.

Kudlyak (2014) studies the cyclical movements of the present value of the wage paid to newly hired workers, $W$. She finds that it falls in recessions and remains persistently low in later years for workers who remain in the same job during the recovery. She concludes that the job creation equation in the basic model is missing an element that makes it unprofitable to hire in recessions, but this element is not wages because labor in recessions is cheap.

A persistent decline in the recruiting cost $\kappa$ is another potential source of recovery in labor-market tightness in recoveries, but it has been studied only in the context of endogenous sources that we will discuss shortly.

Our conclusion is that a gradual improvement in the value of the marginal revenue product of labor that results in a realistic upward trend in $P - W$ is a candidate explanation for the gradual decline in unemployment. However, the empirical support for that mechanism is weak. In particular, there is no systematic improvement in productivity above trend in recoveries.

### 7.2 Financial sources of rising $P - W$ in recoveries

Attention has turned in the DMP literature to financial factors as driving forces for unemployment (Hall (2017), Kilic and Wachter (2018), Kehoe, Lopez, Midrigan and Pastorino (2020)). These papers observe that $P - W$ is the discounted value of the future cash flow.
to the employer, along with other discounted flows in more elaborate models, and thus are sensitive to fluctuations in discount rates. These fluctuations are substantial, according to financial principles set forth in Campbell and Shiller (1988) that are widely accepted in financial economics today.

Discount effects operating through $P - W$ will be present if $W$ moves less than $P$, that is, if the present values of wages are somewhat sticky relative to the present value of productivity. Wage stickiness is fully consistent with DMP principles, provided it is not so severe as to dictate a wage outside the bargaining set of the worker and the employer, which would destroy a match despite its joint value to the parties (Hall (2005b)).

Spikes in general financial discounts coincide with spikes in unemployment, and so are logical candidates to be the source of high unemployment from recessions. But declines in discounts are not nearly persistent enough to account for the lengthy recoveries observed in unemployment.

One potential source of persistent financial effects is a crisis-induced cut in the availability of credit, which raises discount rates for credit-dependent firms and thereby cuts $P - W$. As the availability of credit gradually returns back to normal, unemployment also returns to normal. Dromel, Kolakez and Lehmann (2010) pursue this approach to explaining the high persistence of unemployment.

We provide evidence of the persistent influence of credit conditions on unemployment in recoveries. To measure the availability of credit, we use data from the Federal Reserve Board’s Survey of Senior Loan Officers. Respondents in the survey answer in terms of tightening and easing of commercial loan standards. We cumulate these answers using the statistical model in Hall (2011) to form an index of loan availability. The scale of the index is arbitrary. In Figure 24, we scale it to have the same standard deviation as the observed compound driving force $D_t$ and compare the scaled index to $D_t$. The two variables move closely together. The slow relaxation of lending standards matches the slow decline in the driving force $D_t$.

Petrosky-Nadeau (2014) introduces external financing of vacancy costs in frictional credit markets. The easing of financing constraints during an expansion as firms accumulate net worth reduces the opportunity cost for resources allocated to job creation. Agency-related credit frictions endogenously generate persistence in the dynamics of labor-market tightness. Garin (2015) studies the effects of changes in collateral requirements on the cyclical properties of unemployment and job creation. In the model, borrowing limits are linked to the firm’s physical capital stock. Financial frictions arise from an imperfect enforcement contract. Financial frictions in the form of borrowing constraints create a wedge in the job
creation equation as in Petrosky-Nadeau’s paper. To the extent the constraint is binding, the marginal cost of hiring an employee increases.

8 **Endogenous Mechanisms Implying a Slow Downward Glide in Unemployment during Recoveries**

In this section, we consider self-contained mechanisms that could explain the consistent pattern of recoveries in unemployment. The high persistence of economic activity in general, and unemployment in particular, has puzzled macroeconomists for decades. Finding explanations of endogenous fluctuations in the labor market or other markets has been a goal of many generations of researchers. Our particular interest is labor-market mechanisms operating to generate the observed pattern of reliable but slow recovery of the unemployment from the high levels experienced in recessions.

We study the situation immediately after a major shock has left a legacy of high unemployment. In the basic model, tightness is determined by the equation,

$$\theta = \left( \mu \frac{P - W}{\kappa} \right)^2,$$

which excludes any influence of unemployment except through the driving forces. Even if half the labor force is unemployed, jobs are not hard to find, as long as the driving forces are at normal levels. Instead, the volume of vacancies created by employers is at a high
enough level to bring $\theta$, the vacancy/unemployment ratio, to a normal level. The supply of vacancies is perfectly elastic.

With instantaneous response of tightness to restored normal driving forces, unemployment returns to normal fairly quickly. Recoveries are unrealistically speedy. A number of interesting contributions to the DMP literature, mostly recent, alter the model to mimic the high persistence of unemployment.

One appealing notion in the quest for persistence is that the legacy of high unemployment from a recession creates congestion in the labor market—the high levels of vacancies hypothesized by the canonical DMP model are impractical because employers would interfere with each other just as additional cars joining a crowded highway slow down all of the traffic. A similar mechanism derives an increase in marginal vacancy costs from convex adjustment costs—see Fujita and Ramey (2007).

### 8.1 Negative feedback from unemployment to tightness

We consider a class of DMP-type models in which the unemployment rate influences labor-market tightness $\theta$. We will review an extensive literature that deals with this modification of the basic DMP model. In this class, the $\theta$ function is not a horizontal line as in Figure 18 but rather slopes downward in the unemployment-tightness diagram. For clarity, we discuss a model with no exogenous movements of driving forces. The model combines the natural downward glide of unemployment intrinsic to the DMP model with an offsetting resistance from the endogenous influence of the unemployment rate.

The $\theta$ function becomes

$$\theta(u) = \left( \mu_0 \frac{P_0 - W_0}{\kappa_0} \right)^2 \gamma(u),$$

where $\gamma(u)$ is decreasing in $u$. It captures the negative effect of the current level of unemployment on labor-market tightness, $\theta$, arising from feedback effects. We normalize the variables so that $\theta(u)$ itself measures the influence of unemployment on tightness.

The observed relation between unemployment and tightness during the expansion from 2009 to 2020 is shown as the blue line in Figure 25. The fact that the line fits a smooth curve, except for small transitory deviations, supports the hypothesis that a functional relationship exists between $u$ and $\theta$.

One interesting implication of this hypothesis is that the Beveridge curve should fit perfectly in recoveries. In general, the Beveridge curve has loops because vacancies constitute a jump variable controlled directly by tightness while unemployment is a lagging state variable. Making tightness a function of unemployment eliminates the loops, but only during recoveries.
Figure 25: Phase Diagram for the DMP Model with Negative Dependence of tightness on unemployment

As before, the $\dot{u}$ curve in the figure traces out the relation between unemployment and tightness such that the change in unemployment is zero. It is

$$\theta = \left( \frac{1 - u s}{u \mu} \right)^2,$$

(15)

using the values of the parameters $s$ and $\mu$ just discussed.

A recovery involves a gradual movement along the $\theta$ function starting at the lower right and moving toward the stationary point with unemployment $u^*$. Figure 25 shows that recoveries will proceed slowly, because the two curves are close to each other. In October 2009, at the beginning of the recovery, high unemployment discouraged tightness to a small fraction of its normal value. Job creation proceeded only enough to lower unemployment slowly.

Earlier, in subsection 6.4, we calculated the values of the driving forces, matching efficiency $\mu$ and the job value $J$, from their definitions. We found that both variables rose substantially in the two most recent completed recoveries. If there is no endogenous force taming the DMP model’s tendency toward full employment, some other factor is needed to explain slow recoveries.

Though congestion externalities may play a role in understanding unemployment persistence, it is important to recognize that concavity of the empirical matching function captures the congestion externality to the extent it is reflected in the actual level of vacancies.
Progress in this area will either (1) revise the technology to increase the effective concavity in vacancies, or (2) invoke adverse forces that counteract the incipient high level of vacancies early in recoveries. The resulting modified DMP model would have a lower elasticity of supply of vacancies and higher the persistence of unemployment in comparison to the canonical DMP model.

It is clear that the story of the phase diagram only works if the effect embodied in $\gamma$ is reasonably strong—enough to twist the curve clockwise from flat to downward sloping, and lying close to the $\dot{u} = 0$ curve.

### 8.2 Vacancy costs

The cost of maintaining vacancies, $\kappa$, can be a channel of gradual decline in unemployment in recoveries. One idea is that elevated unemployment in a recession raises $\kappa$ by creating congestion. In turn, as discussed earlier, higher $\kappa$ slows the decline in unemployment. We consider mechanisms to describe this feedback from unemployment to tightness as we described earlier, where the $\theta$-equation involves an offset $\gamma(u)$ depending negatively on unemployment.

The effect of these models in extending unemployment persistence operates through the marginal cost of adjustment, which enters the model through the parameter $\kappa$, now reinterpreted as the derivative of a convex adjustment cost function. The analysis in the previous section considers the response to higher values of $\kappa$ in the compound driving force $D = (P - W)/\kappa$.

This alteration of the basic DMP model eliminates the disconnect of unemployment from the determination of tightness. As we noted earlier, in the basic model this property arises from the perfectly elastic supply of vacancies at cost $\kappa$, a fixed parameter of the model.

If expanding vacancies involves increasing marginal cost, making $\kappa$, the marginal cost of maintaining a vacancy, an increasing function of unemployment, the elasticity of supply of vacancies will be finite. An extensive literature on adjustment costs in vacancy creation has pursued this point.

Adding adjustment costs to the vacancy-creation technology would be a natural way to tame the behavior of vacancies at the beginning of recoveries. Fujita and Ramey (2007) is the pioneering analysis of adjustment costs in vacancy creation. Their model makes the marginal cost of creation rise in proportion to the rate that employers raise vacancy creation. The feedback generates higher persistence of unemployment. When unemployment is high and employers are creating vacancies at higher rates, $\kappa$ is high, offsetting some of the incentive to hire and slowing the recovery. Creation costs induce firms to smooth the adjustment of new openings following a shock, leading the stock of vacancies to react sluggishly. Fujita and
Ramey’s modification of an otherwise standard DMP model eliminates the counter-intuitive property of that model, that vacancies are a jump variable that increases by the full amount of the increase in unemployment in a crisis. Their model makes vacancies a state variable obeying an adjustment process. During that process, the decline in unemployment from a high initial level is slower.

Ferraro (2017) confirms the conclusion in Fujita and Ramey (2007) that the perfect elasticity of supply of vacancies assumed in the canonical DMP model is unrealistic. Within its third-moment framework, the paper argues that upward-sloping supply results in more realistic performance of the DMP model, notably in its ability to match slow recoveries. Coles and Kelishomi (2018) test and reject the assumption of the canonical DMP model that tightness is orthogonal to unemployment. They conclude that the vacancy creation process is less than infinitely elastic.

8.3 Recruiting process and externalities

One line of modeling to support the proposition that higher unemployment raises recruiting costs is the following: Employers have a choice between costly screening of applicants or hiring without screening. In normal times, most employers do not screen, because most applicants self-select to be well matched to the jobs being filled. In times of higher unemployment, self-selection breaks down and employers invest in screening prior to negotiating terms with qualified applicants. The effective cost $\kappa$ of maintaining a vacancy rises and the labor market slackens rapidly. As time passes, conditions gradually reduce $\kappa$ (this is the challenging part) and unemployment begins to decline. The process gains momentum as the pool of job-seekers begins to increase its self-selection. Unemployment gradually declines along the path described earlier in this paper.

This mechanism was considered in Hall (1990) and Hall (2005a). The cost of evaluation per hire depends on the fraction of applicants who are qualified for the job. Applicants may be better informed about their qualifications than are employers. If incentives induce self-selection by job-seekers, so that they apply mainly for jobs where they are qualified, friction and thus unemployment will be low. Self-selection is strongest in markets where unemployment is low and jobs are easy to find. Because of this positive feedback, the equilibrium in a market with self-selection is fragile—unemployment is sensitive to its determinants. Self-selection provides a mechanism for amplification of small changes in the determinants of unemployment.

Gautier (2002) focuses on an externality in the labor market which is caused by non-sequential search. The job-offer rate is increasing in the number of applications while the hiring rate is decreasing in the flow of applications per applicant. The externality arise
because screening information is lost when a worker is found unsuitable for a job and the next firm has to spend time screening the applicant again.

Villena-Roldan (2012) builds a model of search equilibrium in the labor market with endogenous recruiting effort involving the employer’s choice of the number of candidates to screen for a given job opening. Although the paper does not focus on our issue, it appears that matching efficiency is lower in an equilibrium with higher unemployment.

Molavi (2018) investigates conditions that lead to an outward shift of the Beveridge curve and thus a decline in matching efficiency. To the extent that a force that shifts unemployment has this effect, his paper may contribute to an understanding of the changes that occur in the labor market during a recovery.

Engbom (2021) takes on the challenge of this paper, starting from the observation that unemployed job-seekers bombard recruiting firm with vastly more applications than do on-the-job applicants. As a result, they apply for many jobs that are not actually good fits, thus imposing higher recruiting costs on employers. Higher cost results in lower labor-market tightness for standard DMP reasons. He provides evidence that recruiting effort rises in times of higher unemployment.

Fishman, Parker and Straub (2020) develop a dynamic model of credit markets in which lending standards and the quality of potential borrowers are endogenous. Lending standards set privately by the banks have negative externalities and are dynamic strategic complements—tighter screening worsens the future pool of borrowers for all banks and increases their incentives to screen in the future. Lending standards can amplify and prolong temporary downturns, affecting lending volume, credit spreads, and default rates. In the model, when markets recover, they may do so only slowly, a phenomenon the authors call “slow thawing.” This line of thought may apply to labor markets.

Lockwood (1991) develops a setup where employers may administer a test. Then employers also consider unemployment duration as informative about how many times the job-seeker has flunked previous tests. When unemployment is higher, this problem worsens, creating a congestion externality.

8.4 Composition effects

Eeckhout and Lindenlaub (2019) observe that, in times of high unemployment, the composition of job-seekers shifts toward the unemployed and away from on-the-job searchers. The latter are not included in unemployment. Thus unemployment is not a good measure of the flow of matches—calculations of matching efficiency based on the unemployment rate overstate the growth of efficiency during recoveries.
Hall and Schulhofer-Wohl (2018) study job-finding rates and match efficiency in CPS data broken down by multiple categories based on the personal circumstances of working-age individuals. These include individuals who are unemployed for various reasons, those currently employed, and those out of the labor force who are and are not interested in working. These categories are further broken down by the duration of unemployment to date in the cases of the unemployed. All categories have positive job-finding rates, ranging from high values for most of the unemployed to quite low values for those not interested in working. They calculate a measure of matching efficiency using their findings. It has a smooth trend but does not track the cycle. Our finding displayed in Figure 22 of strong cyclical shifts in efficiency reflects major cyclical changes in the *composition* of unemployment.

See also Sahin, Song, Topa and Violante (2014) and Hornstein and Kudlyak (2016).

Mercan, Schoefer and Sedlacek (2020) propose a model in which newly hired workers are imperfect substitutes for seasoned workers. In their model, a greater share of the unemployed among the potential new hires in the recessions discourages job creation and helps explain the persistence of aggregate unemployment following an adverse shock.

### 8.5 Scarring effects

Other models have incorporated the property that feedback from unemployment causes $P - W$ to decline when unemployment is high. One way for such a feedback is when $P$ declines more than $W$ in response to higher unemployment. Ljungqvist and Sargent (1998) propose a model in which workers accumulate skills on the job and lose skills during unemployment, while their non-employment option remains unchanged. In turbulent economic times, the loss of skills is faster and the decline in productivity is greater.

### 8.6 The separation rate

The separation rate $s$ is another parameter of the DMP model that could contribute to the explanation of the slow recovery of unemployment. An elevated separation rate shifts the stationary locus in the DMP phase diagram to the right. A gradual decline in the separation rate results in a gradual decline in unemployment. Our data on the separation rate support that account of unemployment persistence. Figure 23 shows that the separation rate has a general downward trend with spikes in recessions and slow recoveries back to trend. Research in the VAR framework provides evidence for the importance of job loss in understanding unemployment dynamics. See Fujita (2011), Barnichon (2012), Fujita and Ramey (2012), and Portugal and Rua (2020).
We believe that the separation rate is, apart from trend, an endogenous variable that reflects some of the same forces that keep unemployment on its slow glide path. Its own path is consistent with that hypothesis. In complete contrast to the sharp and short spike in layoffs, shown in Figure 2, separations follow the same kind of glide path as unemployment. Churn set off by a recession gradually recedes, and unemployment, separations, and other measures of labor-market activity calm down in parallel.

9 Other Forces Operating during Recoveries

The DMP model provides a disciplined framework for studying the issues considered in this paper. But a great deal of business-cycle thinking occurs outside the DMP framework. In this section we examine the behavior of policy instruments and other potential driving forces without trying to determine how they might operate through the DMP model.

Next we take a look at a variety of macro variables that may be involved in recoveries. These are policy instruments—government spending and monetary policy—and influences that might be considered exogenous determinants—productivity, labor-force growth, and the stock market.

We use the NBER business-cycle chronology, so that our timing results are measured over the general business cycle, rather than a cycle pertaining specifically to unemployment.

9.1 Fiscal and monetary policy

**Government purchases.** Figure 26 displays consolidated government purchases of goods and services divided by the CBO’s potential GDP series. The dates of peaks in the business cycle appear along the bottom—not the peaks in the purchases series itself. Essentially all macroeconomic models agree that an increase in government purchases stimulates output. The figure shows that purchases in the first recovery, 1949 through 1953, grew rapidly because of the Korean War. The Reagan military buildup in the 1980s also accounted for rising purchases relative to potential GDP in that recovery—in all other recoveries, even the one in the 1960s containing the Vietnam war, purchases failed to keep up with potential GDP. The conclusion with respect to those, notably including the most recent recovery, is that fiscal policy taking the form of deliberate expansion of purchases—such as the American Recovery and Reinvestment Act—provided stimulus when the economy was weak. As the economy recovered, the stimulus was withdrawn.

**Government transfers.** The US has large and effective countercyclical government transfer programs and practices. Figure 27 shows the history of dollar benefits in terms of our unemployment recovery chronology. We standardize the data by dividing by nominal
disposable income. Some of the countercyclical pattern arises from automatic stabilizers—programs that enroll more dependents in bad times—and some from discretionary expansion of programs and creation of new ones—such as extending unemployment insurance benefits to cover more weeks.

The figure shows that there is a good deal of heterogeneity across the recoveries. Transfers declined remarkably in the first recovery, starting in 1949. In the next four recoveries, transfers grew relative to disposable income. In four of the recent five recoveries, transfers declined.

**Monetary policy.** The central instrument of monetary policy in the US is the Federal Reserve’s policy interest rate. The standard way to state its effect as an instrument is to define it as the margin of the economy’s natural or equilibrium short interest rate over the policy rate. To expand, the Fed depresses the policy rate and increases the margin. And to contract, the Fed raises the policy rate above the natural rate to drive the margin negative. Laubach and Williams (2003) is a widely used estimate of the natural short rate.

Figure 28 shows the expansionary margin of interest-rate policy, according to Laubach and Williams. The Fed has chosen net expansion in four expansions and net contraction in two. In the recovery from the 2007-09 recession, the Fed has chosen substantial expansion, almost as much as in the recovery of second half of the 1970s. Oddly, the late 1970s were
a period of high and rising inflation, so the Fed was failing in its duty to lean against the wind.

As with the other policy instruments, we find heterogeneity in the setting of the Fed’s interest-rate margin during the recoveries of the past 70 years.

9.2 Other forces during recoveries

Financial discounts. Forces other than macroeconomic policy may influence unemployment declines during recoveries. For example, a recent literature has described a relation between financial discounts and unemployment. See Hall (2017) in the context of the aggregate labor market and Kilic and Wachter (2018) and Kehoe et al. (2020) in general equilibrium. These papers consider DMP-type models of unemployment and events that alter economy-wide discount rates, thus changing the job-value, which is the present value of the contribution of a newly hired worker net of the wage paid to the worker. Discounts sometimes jump upward almost discontinuously, as they did immediately after the Lehman bankruptcy in 2008. The job value represents the incentive to recruiting. When it declines, the labor market slackens and unemployment rises. In the recovery phase, falling discounts raise the job value and unemployment falls.

According to principles of modern finance elucidated in Campbell and Shiller (1988), discount rates for risky future cash payouts are equal to the expected rates of returns associated
with those payouts. In a recovery, the stock market rises, the price/dividend ratio rises, and expected rates of return decline. According to the literature linking financial events to the labor market, unemployment declines back to normal. Figure 29 shows the history of the ratio for recoveries since 1949. The ratio rose dramatically during the recovery of the 1990s. It fell substantially during the financial crisis in 2008 and 2009, but recovered during 2010, when unemployment was still rising. Its relation to the business-cycle chronology in earlier years is less apparent. A rising price/dividend ratio is sometimes important for a recovery, but does not explain the reliability of US business-cycle recoveries.

**Productivity growth.** Another aggregate influence of unquestioned importance for GDP growth is productivity growth. If the topic of this paper were real GDP growth in recoveries, productivity would receive top billing. But the relation of productivity growth to the gradual rise of economic activity in recoveries is ambiguous and may well be small. Figure 30 shows that productivity level. The productivity growth tended to be high in recoveries through the 1980s, had a small comeback in the recovery starting in 2003, and had a spectacular shortfall in the recovery from the 2007-09 recession. Overall, productivity growth tended to be irregular in recoveries.

**Variations in labor-force growth.** The DMP model of Mortensen and Pissarides (1994) has a constant labor force. Extensions to endogenous participation may involve positive or negative co-movements of participation and unemployment. Figure 31 shows that the
Figure 29: Price/Dividend Ratio of the S&P 500 during Business-Cycle Recoveries

Figure 30: Total Factor Productivity during Business-Cycle Recoveries, Quarterly
participation rate grew during the years up to 1990 when the rising rate for women was a key factor for overall participation (to achieve a basic adjustment for demographic influences, the data refer to ages 25 through 54). In the recovery from the 2007-09 recession, participation was essentially unchanged. In the recovery from the 2020 recession, participation declined.

9.3 Discussion of policies and other forces operating during recoveries

Based on this evidence, we conclude that the economy includes a strong internal force toward recovery that operates apart from policy instruments and from financial developments or productivity growth. Policymakers understand this point and withdraw expansionary policies as the internal force does its job.

We should be clear that optimal policy that resulted in uniform growth of economic activity might be quite irregular as it fights off disturbances, so the irregularity of instruments is not conclusive evidence of the irrelevance of policy. That is, the evidence of stable outcomes and unstable policy instruments is also consistent with the view that policymakers understand the workings of the economy well and deploy the instruments to deliver stable outcomes.

Our tentative conclusion that policy has little impact on unemployment during recoveries still leaves room for effective policy to prevent or moderate recessions.
10  Conclusions: The Sources of the Slow, Stable Downward Glide of Unemployment during Recoveries

We state our conclusions within the framework of the DMP model. That framework hypothesises that there is a powerful economic incentive in normal times to expand employment and drive down unemployment. Early in the paper we established the factual background for our investigation. Occasionally the US economy undergoes a crisis that results in a burst of job loss. In the ensuing years, the recovery process described in the DMP model proceeds, smoothly but slowly. We quantified job loss and the process of finding new jobs. We noted that the process has the character described in the simple DMP model but has some important complexities—even though the typical job-seeker finds a job in a month or two, the net rate of job-finding is much lower, because the early jobs tend to be quite short. But even taking that finding fully into account, we reach an important conclusion: the unemployment paths of the original job-losers are neither as high to begin with nor as persistent as the path of elevated unemployment in the wake of the crisis.

That conclusion implies that some mechanism generates a large volume of additional unemployment beyond the initial job losers from the crisis. One possibility we consider is that the incentive to create jobs in the DMP framework is diminished for an extended period after a crisis by forces outside the labor market, possibly a continuation of the adverse force that caused the crisis. We conclude that it is difficult to pin down such a force. In the original DMP model, that force was a persistent shortfall in productivity, but we observe, along with most students in the DMP school today, that a collapse of productivity in a crisis followed by gradual restoration of productivity holds no water whatsoever. A more recent attempt to affiliate the DMP model with a financial driving force is successful in explaining the spike in unemployment in crises, but not in explaining the slow rates of recoveries.

We examined some data on the availability of credit from banks that suggests a financial influence with on unemployment with persistence comparable to that of unemployment. We also studied various measures of churn in the labor market, including the separation rate from jobs, which is one of the driving forces of the DMP model. Churn declines along with unemployment over the duration of a recovery.

We studied the hypothesis that unemployment is contagious—times of high unemployment in the aftermath of a crisis are times when job-seekers find it hard to find jobs. It seems logical that jobs are hard to find when unemployment is high, but that is not the way that the DMP model works. With its emphasis on the incentives for job creation, the DMP model holds that the determinant of tightness in the labor market is the level of that incentive. If employers find it potentially profitable, they will exert the same effort to lay
on new workers if the unemployment rate is 10 percent or 4 percent—vacancy creation is infinitely elastic. Contagion may arise from the congestion that occurs in the recruiting process when employers are flooded with applicants. Or contagion may involve changes in the equilibrium search and recruiting strategies of job-seekers and employers that impede matching and lower the efficiency of the matching process. These modifications of the DMP model lower the elasticity of vacancy creation and slow down the rate of recovery of unemployment.

Invoking adjustment costs in recruiting effort is another promising alteration to the DMP model. It would tame the model’s enthusiasm for hiring when unemployment is high but incentives are only normal. In the case of investment in physical capital, when the payoff to capital rises, firms do not instantly buy all the additional capital merited by the higher return. Models of adjustment costs for capital typically make the flow cost of capital installation rise with the volume of new capital installed—adjustment-cost functions are often quadratic. In the presence of adjustment costs, the investment process spreads over time. The same moderating role of adjustment costs applies to investing in workers.

We noted that part of the high level of unemployment soon after a crisis reflects a change in the composition of unemployment toward individuals with naturally lower job-finding rates. This is a source of lower matching efficiency, a decline in one of the DMP model’s driving forces. A related phenomenon is the lower incidence of on-the-job search when unemployment is high. Again, this is a source of lower matching efficiency.

The DMP model imposes serious restrictions on the forces that govern the speed of recovery of the labor market after a crisis creates a burst of unemployment. Most of our investigation operates within those restrictions—we take the DMP model seriously. But we do take a preliminary look at the relation between the behavior of unemployment during recoveries and other macro variables: government purchases and transfer payments, monetary policy, productivity growth, financial discounts, and labor-force participation rates. We do not spot any pattern of co-movement of these variables with unemployment. Our tentative conclusions do not rise to the level of firmly established causal inference. Much more remains to be done.
References


53


