What's Driving Wage Inequality?

Aaron Steelman and John A. Weinberg

M ost of the time, we assess an economy's performance using broad aggregate measures of output and wealth. In this regard, the United States is doing quite well. It is the richest country in the world. U.S. gross domestic product exceeded $11 trillion last year—roughly $38,000 per capita. And despite the slowdown associated with the 2001 recession, the economy has expanded at an average annual rate of more than 3 percent over the past 10 years. The way people actually feel about the economy's performance is shaped by their individual experiences, however, and here there is always great diversity. Indeed, there remains substantial anxiety about the direction the economy is heading, especially in regard to the growing disparity in income. The gap in real wage rates between those at the higher end of the distribution and those at the lower end has been widening for some time. In addition, the real wages of workers at the lowest part of the distribution were stagnant or falling during much of this extended period of growing wage inequality.

This essay will explain why wage inequality has been increasing in the United States; in doing so, we will draw upon the scholarly literature, including work done by Richmond Fed economist Andreas Hornstein with Per Krusell of Princeton University and Giovanni Violante of New York University. We also will discuss the associated policy implications—that is, what can be done to better assure that all Americans have the opportunity to secure well-paying jobs, as well as which policies may hinder that goal.

Overall, we will argue that technical innovation has significantly affected the wage distribution in the United States. But the direction of that effect has

not been uniform. In the early part of the twentieth century, various technical innovations had the effect of compressing the wage structure. Since the 1970s, however, technical innovation—particularly the introduction and widespread use of information technology—has produced wage dispersion.

Another force to which many have attributed recent labor market developments is globalization. We conclude that international trade and immigration, while significant trends, are not by themselves the primary force behind growing wage inequality. To some extent, globalization is itself a result of advances in information technology, which allow the production of goods and services to take place over a broader geographic area.

As for public policy, research suggests that increased emphasis on education is a sound response to recent trends in wage inequality, particularly education early in life and programs focusing on general, broadly applicable skills. Early skill acquisition yields rewards over a relatively long period of time because individuals can recoup their investment in human capital throughout their working lives. In addition, such training tends to build on itself: acquiring skills early in life makes it easier to acquire additional skills later in life. In contrast, policies that would aim to slow the growth in wage inequality by imposing barriers to globalization, such as trade restrictions, would likely do little to achieve their intended goal, while lowering aggregate income and overall social welfare.

Before discussing why wage inequality has been growing and the steps policymakers may wish to consider in response, it is necessary to look at the facts. In the next section, we present data on wage inequality from the early twentieth century to the present.

1.   THE FACTS

Most economists agree that wage inequality has been increasing in the United States recently.[1] But this has not always been so. Wage inequality was large during the first part of the twentieth century, decreased during the middle part of the century, and accelerated again toward the end of the century.

During the early part of the twentieth century, several factors contributed to a decline in the demand for less-skilled workers. For instance, the widespread introduction of electricity and new hoisting equipment in the 1910s greatly reduced the need for common laborers who moved goods to and within factories.[2] The lower demand for these workers' services put downward pressure on their wages. At the same time, the rise of large businesses increased the demand for the relatively small subset of workers with higher education to fill

---

[1] For an exception, see Lerman (1997).
[2] Goldin and Katz (1999, 9).

managerial roles, thus driving up their wages. As a result, wage inequality grew during the first quarter of the twentieth century.

By the 1940s, wage structures began to change significantly, however, so much in fact that Claudia Goldin and Robert Margo have called this period "The Great Compression," describing the general decline in wage inequality.[3] On the supply side, the once small number of college graduates began to face increased competition, as thousands of American military personnel came back from World War II and took advantage of the GI Bill. This influx of newly minted graduates most likely helped depress the relative earnings of college-educated workers. In addition, the quality of education at the high school level became less variable during this period, meaning that the skill differentials between high school graduates in different parts of the country probably decreased, thus reducing the disparity in wage rates among this group of workers.

On the demand side, more low-skilled labor was needed in the nation's industrial centers to produce goods for the war effort, therefore driving up the relative wages of these workers. In addition, government intervention through the National War Labor Board almost certainly contributed to the compression of the wage structure.[4]

It is interesting to note that there is also evidence of wage compression in the United Kingdom during the Industrial Revolution of the eighteenth and nineteenth centuries. Goods that were once produced by artisans in relatively small numbers over relatively long periods of time were produced in factories following industrialization.[5] This meant that more-skilled workers were replaced by less-skilled workers, who because of the introduction of interchangeable parts and other production techniques could perform their tasks efficiently with little training. The demand for low-skilled workers, then, increased during this period, demonstrating that not all technological innovations are necessarily "skill-biased." Some, in fact, have been "skill-replacing."

That brings us to the last half of the twentieth century. In particular, we will focus on the period from 1970 onward. As stated earlier, this has been a period of growing wage inequality. Consider the following observations.[6]

---

[3] Goldin and Margo (1992).

[4] The National War Labor Board was created in 1942 in an effort to stabilize wages during World War II. According to two authors who worked at the agency, "no changes in wage rates could be made except upon approval of the National War Labor Board; and ... the Board could approve wage increases only on four narrowly circumscribed grounds, and wage decreases on only two grounds." See Henig and Unterberger (1945, 319–20).

[5] For more on the introduction of new technology in England during the Industrial Revolution, see Mokyr (1994).

[6] These observations are taken from Hornstein, Krusell, and Violante (2004), which surveys empirical work up to 1995. Recently, Eckstein and Nagypál (2004) and Autor, Katz, and Kearney (2004) have updated some of these observations. Instances in which the more recent observations differ from the older observations are noted in the text.

- The 90-10 weekly wage ratio, which compares the wages of workers at the 90th and 10th percentiles of the wage distribution, rose from 1.20 to 1.55 for males and from 1.05 to 1.40 for females from 1965 to 1995. Similar growth in inequality was found elsewhere in the wage distribution, though dispersion in the lower wage groups (for instance, the 50-10 ratio) seems to have stabilized recently.

- Average and median real wages have changed little since the mid-1970s. But real wages in the bottom 10 percent of the wage distribution fell sharply during much of this period before experiencing modest growth recently. Meanwhile, the real wages of those at the top of the distribution, especially the top 1 percent, have risen sharply.

- The returns gained from education fell in the 1970s, but have increased since. The college wage premium—defined as the ratio between the average weekly wage of a college graduate and a worker with a high school diploma or less—was 1.35 in 1975, 1.5 in 1985, and 1.7 in 1995.

- The returns from experience also grew in the 1970s and the 1980s but flattened in the 1990s. For instance, the ratio of weekly wages between workers with 25 years of experience and workers with five years of experience increased from 1.3 in 1970 to 1.5 in 1995.

- The returns from white-collar occupations relative to blue-collar occupations increased by about 20 percent from 1970 to 1995.

- Inequality across race and gender has declined since 1970. The black-white differential and the male-female differential have both dropped. Also, labor force participation of women increased dramatically during this period.

The last three points all involve "between-group" comparisons—that is, comparisons of workers classified by observable characteristics, such as education, experience, occupation, race, and gender. But it is also true that wage inequality "within groups"—that is, among workers with similar education or experience, for instance—has risen. This trend seems to have started about a decade prior to the trend of increasing returns from college education.[7] Looking abroad, recent trends in wage inequality in the United Kingdom tend to resemble those in the United States. Things in continental Europe are quite different, though. There has been almost no increase in wage inequality there. Indeed, wage inequality has even declined in Belgium, Germany, and Norway.

---

[7] Juhn, Murphy, and Pierce (1993, 412).

2. THE ARGUMENT

What is driving the increasing disparity in wages in the United States? The evidence strongly suggests that there has been skill-biased technical change that has benefited the more-skilled workers over the past 30 years. By skill-biased change, we mean advancements in technology that have boosted the productivity of skilled labor relative to that of unskilled labor.

To determine why this is the case, it is important to understand that the relative wages of workers at different skill levels are determined by the relative supply of and demand for those types of workers; that supply is determined by the relative number of more-skilled and less-skilled workers; and that demand for those workers' labor is determined by the current state of technology, which in turn largely determines the productivity of different types of labor.

At first, this explanation may appear to fit awkwardly with the facts. After all, the relative supply of more-skilled workers, measured as a fraction of the workers with a college education, has risen sharply during this period. Wouldn't this increased supply tend to depress wages, as seemed to happen at mid-century? Standard theory would suggest yes: with a given demand, more supply of a good would tend to drive down its relative price. And for a while this seems to have been the case with skilled labor. During the 1970s, the number of college graduates rose sharply and effectively flooded the market, driving down the returns gained from education. But by the 1980s, more-skilled workers were able to command a wage premium.

What accounts for the change? In large measure, the development of new technology. In particular, information technology, which began to make its way into the workplace in the 1970s but did not become widespread until the 1980s, the same time as the returns from skill began to increase. What is it about information or computer technology that increases the demand for skilled workers? According to David Autor, Frank Levy, and Richard Murnane, two mechanisms—substitution and complementarity—are at work:

> Computer technology substitutes for workers in performing routine tasks that can be readily described with programmed rules, while complementing workers in executing nonroutine tasks demanding flexibility, creativity, generalized problem-solving capabilities, and complex communications. As the price of computer capital fell precipitously in recent decades, these two mechanisms—substitution and complementarity—have raised relative demand for workers who hold a comparative advantage in nonroutine tasks, typically college-educated workers.[8]

Autor, Levy, and Murnane conclude that information technology can explain between 60 and 90 percent of the estimated increase in relative demand

---

[8] Autor, Levy, and Murnane (2003, 1322).

for college-educated workers from 1970 to 1998. So while the relative supply of more-skilled workers certainly increased during this period—which, all else being equal, would have tended to depress the relative wages of such workers—the demand for such labor increased even more because of technical change.

Consider a few examples that may help to illustrate their point. Advances in manufacturing, such as the introduction of computer-controlled machinery, have often meant fewer workers on the factory floor with those remaining needing a higher level of skill to operate the increasingly sophisticated equipment. A similar process is at work in the division of labor between architects and draftsmen. Before the advent of computer-aided design—or "CAD"—a draftsman would create and revise plans under the guidance of an architect. With CAD, however, the architect can easily generate and manipulate plans on the computer, resulting in the employment of fewer draftsmen, while boosting the productivity of the overall design process.

Some economists have suggested that the increasing supply of skilled workers may have actually induced the development and implementation of new technologies that require higher levels of skills. In short, as Daron Acemoglu has argued, "When developing skill-biased techniques is more profitable, new technology will tend to be skill-biased."[9] Conversely, when developing skill-replacing techniques is more profitable, new technology will tend to be skill-replacing. This, arguably, is what happened in England during the Industrial Revolution. The migration of large numbers of less-skilled workers to the English cities from rural areas and Ireland made the implementation of skill-replacing technologies profitable. "So, it may be precisely the differential changes in the relative supply of skilled and unskilled workers that explain both the presence of skill-replacing technical change in the nineteenth century and skill-biased technical change during the twentieth century."[10]

Thus, overall, the best explanation for the increase in wage inequality appears to be skill-biased technical change. But there are some potential challenges to this theory.

3.   THE CHALLENGERS

Not all economists are persuaded that increasing returns from skill were the principal driver of wage inequality during the 1970s. Some have offered competing explanations, many of which are centered around institutional change.[11] One explanation, for example, is the erosion of the real value of the minimum wage and the decline in unionization in the United States. Other theories focus

---

[9] Acemoglu (2002, 9).

[10] Ibid., p. 12. Also, see Acemoglu (1998).

[11] See, for instance, Card and DiNardo (2002).

on globalization—specifically, increased trade with less-developed countries (LDCs) and immigration of less-skilled workers to the United States. Finally, some point to evidence from other countries. If skill-biased technical change is causing growing wage inequality in the United States, they ask, why isn't wage inequality also growing rapidly in Western Europe since all developed countries have access to basically the same technology? We will address those issues in turn.

The nominal value of the minimum wage remained constant throughout much of the 1980s, meaning that as prices rose its real value dropped. Because the minimum wage may be expected to raise the wages of low-paid workers, the decline in its real value could be responsible for increased wage inequality.[12] There are three problems with this hypothesis, though. First, the number of U.S. workers—especially male workers—affected by the minimum wage is quite small, less than 10 percent of all workers between the ages of 18 and 65. Second, the erosion in the real value of the minimum wage occurred in the 1980s, while the general trend of rising wage inequality began in the 1970s. One would expect the two to coincide more closely if the decline in the real value of the minimum wage were indeed a significant factor. Third, a large share of the increase in wage inequality is due to rapid gains by workers at the top of the wage distribution. For these people, the minimum wage is not a binding constraint.

Timing is also a problem in theories that focus on declining unionization.[13] The 1950s, as we have discussed, was a time of wage compression, not growing wage inequality. Yet it was during this decade that unionization began its steady decline. To be sure, the decline of unionization in the private sector picked up pace during the 1970s and 1980s. But at the same time, the public sector workforce became increasingly unionized, compensating for some of the loss in the private sector. In addition, wage inequality has increased quite rapidly in some sectors of the economy that were never highly unionized, such as the legal and medical professions.

There is, however, some evidence that technical change may have been partially responsible for the decline in unionization since the 1950s.[14] Such a decline could have caused the real wages of low-skilled workers to fall (a point that we will return to in the next section), but its effect on increasing wage inequality would have been only indirect, with technical change starting the whole process.

Popular opinion often attributes increased trade with LDCs as the principal cause of increasing wage inequality in the United States—an explanation that

---

[12] Lee (1999) argues that this has, in fact, occurred.

[13] For a recent paper that argues that there is a significant relationship between unionization and wage inequality, see Card, Lemieux, and Riddell (2003).

[14] See Acemoglu, Aghion, and Violante (2001).

some economists have argued is consistent with the data. Indeed, standard trade theory, based on the principle of comparative advantage, would seem to predict just that. Since LDCs have relatively large numbers of unskilled workers, an increase in trade would act like an increase in the relative supply of unskilled workers in the United States, thus potentially increasing wage inequality. And trade between the United States and the developing world has indeed increased substantially during the past 30 years, the period during which wage inequality has been increasing.

The relative price of skill-intensive goods has not increased over the period of rising inequality, however, as one would have expected if trade were a significant factor in wage dispersion. Perhaps more telling, the total volume of trade with LDCs is arguably too small to have had a significant effect on U.S. wages. The effects of trade flows on "relative skill supplies have not been substantial enough to account for more than a small proportion of the overall widening of the wage structure over the past 15 years and have played only a modest role in the expansion of the college-high school wage differential in the United States," conclude George Borjas, Richard Freeman, and Lawrence Katz.[15]

As for immigration, the total number of newcomers to the United States during the period under review also is probably too small to have had a large effect on the wage structure. For instance, during the 1970s, immigration added 2 million new workers to the U.S. labor force. But because of the baby boom and the increased participation of women in the workplace, roughly 20 million new native workers also entered the labor force during that period. In addition, even during the 1980s, a period of relatively high immigration, the immigrant share of the total labor supply increased by only 1 percentage point, from 7 to 8 percent. "These magnitudes can be taken to mean that immigration is unlikely to have large effects on the overall distribution of wages," concludes Robert Topel.[16]

Finally, some have argued that if technical change is a significant cause of wage inequality, then it ought to have affected the wage structure in Western Europe in the same way that it has in the United States since those countries have access to much the same technology and arguably employ it in similar ways to American firms. But, as we know, wage inequality has not increased as rapidly in Western Europe as it has in the United States. Does this cause significant problems for the skill-biased technical change explanation of wage inequality? Some have suggested so. We think otherwise, however. The observations from Western Europe can be explained by factors that do not contradict the skill-biased technical change argument.

---

[15] Borjas, Freeman, and Katz (1997, 67).

[16] Topel (1997, 62).

As many commentators have noted, Western Europe has significantly less flexible labor market policies than the United States, including more comprehensive employment protection, longer and more generous unemployment benefits, and greater restrictions on wage bargaining. Those policies likely have had the effect of compressing wages. Thus, while similar technical change may have been introduced at roughly the same time in the United States and Europe, different labor market policies have resulted in different effects on the respective wage structures.[17]

In addition, Europe's labor market policies combined with rapid technological change arguably have led to greater unemployment. In the 1960s, the United States and Europe had roughly the same unemployment rates. Since then, Europe's labor market policies have not changed substantially—those policies have been restrictive for many decades—but its unemployment rate has risen sharply. Why?

Strict employment-protection laws make it difficult for companies to terminate workers in Europe. But over time some workers will leave voluntarily, perhaps encouraged by generous social-welfare benefits. Those workers' skills become dated quickly as technology changes, just as they do for unemployed workers in the United States. But the principal difference is that the strict European employment-protection laws that made those same workers difficult to terminate in the first place also have the effect of keeping them out of the workforce longer than they would have been otherwise. Employers, knowing that all new hires are possibly lifelong employees, will look very carefully for a good match. Those workers whose skills are not up-to-date will have difficulty finding new employment. And the longer they are out of work, the more difficulty they will have, because multiple generations of technology will have been introduced and replaced during their absence from the workforce. Also, the generous welfare benefits those workers receive reduce their incentives to acquire new skills on their own.

In the United States, where it is easier to terminate workers, employers do not have to be as careful when hiring new employees. The cost of taking a chance on a worker whose skills may be somewhat dated is potentially much smaller than in Europe. As a result, the U.S. unemployment rate has not risen steadily over the past 30 years, as it has in most European states.[18]

4. THE PROBLEMS

We have argued that the most compelling single explanation for the rise of wage inequality in the United States since the 1970s has been skill-biased technical change. In addition, we have argued that other proposed explanations—such

---

[17] See Krugman (1994).

[18] For a complementary explanation, see Ljungqvist and Sargent (1998).

as institutional change and globalization—do not appear very persuasive. Yet there remain two unresolved issues.

First, as we previously noted, the growth of wage inequality within groups, sometimes referred to as "residual inequality," is quite large and may not be adequately explained by skill-biased technical change alone. Second, and also mentioned earlier, real wages for those at the lowest end of the distribution declined during much of the last 30 years. Yet, as Acemoglu has argued, it is unclear how "sustained technological change can be associated with *an extended period* of falling wages of low-skill workers."[19] How can these developments be explained?

Perhaps the most compelling explanation for the increase in residual inequality is that there are unmeasured differences in the skills among workers within groups. Consider, for example, two economists that have nearly identical profiles: both are 50-year-old, white males; hold graduate degrees from similar institutions; and have worked as university professors for 20 years. To an outside observer, it is impossible to distinguish between the two workers. But to their colleagues and students, there may be very substantial differences. One economist simply may have more natural talent than the other, producing innovative research across a number of fields. Or he may be a more gifted teacher who inspires students in the classroom. In either case, he is a more valuable worker than his counterpart and consequently may receive a higher wage. We should not be surprised by such a wage differential, but according to our measures of worker characteristics, both economists fall into the same group—thus leading to an increase in residual inequality. Skill-biased technical change increases the premium paid to skilled workers, even if skills are not well-measured by such characteristics as education or experience.

Also, rising residual wage inequality may be possible even without unmeasured skill differences. One possible explanation of this phenomenon involves the role of vintage capital. Close examination of the data suggests that the pace of technological advancement has been accelerating since the mid-1970s. Yet different firms have adopted new technologies at different times and at different levels; that is, firms employ technologies of different vintages. This has important implications for the wage structure. In a model that includes labor market frictions—meaning that the labor market is not fully competitive because, for instance, it is costly to switch jobs—workers with the same skills can be expected to earn different wages. More specifically, their wages will increase as the productivity of the technology with which they are working increases. As a result, it is plausible that technological acceleration may increase wage dispersion within groups, since with more rapid technical change you have more vintages of technology in operation simultaneously.

---

[19] Acemoglu (2002, 13).

But what about the drop in real wages of less-skilled workers? In a world of relatively slow technical change, many skills are easily transferable. Workers can move from one company to another with little trouble adapting to the machinery at their new firm. In a world of rapid and accelerating technological change, however, such moves are more difficult since fewer skills are transferable. Upon separation—that is, when workers leave a firm—those workers can expect to suffer wage losses. This scenario is especially true of workers who have been using the oldest technology, because they find that the skills they have acquired through experience are even *more* outdated than those of workers in similar industries who have been exposed to more modern technology. Thus, accelerating technological change may help us explain both the rise in residual inequality and the decline in real wages at the bottom of the distribution.[20]

It is important to note, though, that such conclusions are only tentative. Whereas there seems to be overwhelming evidence and an emerging consensus about the role of skill-biased technical change on the wage structure, there remains a good deal of uncertainty about the cause(s) of residual inequality and the declining real wages of less-skilled workers.

5.   IMPLICATIONS FOR PUBLIC POLICY

What lessons should policymakers draw from our discussion of the causes of wage inequality in the United States? We might start with a general principle that is often associated with the medical profession but is applicable to public policy as well: first, do no harm. There is understandably a great deal of anxiety among the public about the changing nature of the American economy. Those forces which create economic growth for us all, also cause disruptions for some.[21] As Joseph Schumpeter famously noted, capitalism is characterized by "the perennial gale of creative destruction."[22] And to many people, that gale—at least for the moment—is associated with globalization.

Yet, as we have argued, increased trade with LDCs and immigration from abroad likely have had little effect on wage inequality, while almost certainly adding to the strength and vitality of the American economy.[23] Efforts to slow the growth of foreign goods or labor coming to our shores would be costly to Americans as a whole, as well as to those people who seem to be hurt by globalization at the present. As Jeffrey Sachs and Howard Shatz have written,

---

[20] This section draws on Violante (2002).

[21] Fears about the effect of technical change on the job market—in particular, the belief that technical innovation is a net destroyer of jobs—is not new. David Ricardo and other classical economists addressed the issue. See Humphrey (2004).

[22] Schumpeter (1942).

[23] See Burtless, Lawrence, Litan, and Shapiro (1998) for a discussion of the benefits of open trade. See Simon (1999) for a discussion of the benefits of liberal immigration policies.

"U.S. labor market experience . . . teaches that the labor force will respond to the premium on education by increasing the investment in education, thereby narrowing the gap in inequality in the future."[24] Insofar as barriers aimed to slow globalization dampen the incentive to build skills, those barriers will tend to perpetuate wage inequality.

In addition, we should be wary of proposals to extend the duration of or expand the generosity of unemployment insurance benefits to those workers who have lost their jobs due to technical change. Such proposals would tend to increase the time that displaced workers remain unemployed. Instead, we ought to encourage those workers to reenter the labor force as quickly as possible. The problem, of course, is that the jobs that such workers will be able to secure will likely pay significantly less than their former positions. "Workers not only lose income when they are unemployed, but many often suffer a drop in their earnings after finding new jobs. Older workers—who tend to be less flexible adapting to new production techniques or who lack the educational background to transfer to well-paid service economy jobs—bear the greatest losses," write Lori Kletzer and Robert Litan.[25]

An alternative way to assist displaced workers may be a simple transfer program that subsidizes their wages upon reemployment.[26] This policy would boost recipients' incomes, while allowing them to allocate their financial resources toward the mix of training opportunities and general consumption they deem most beneficial. Such a program would certainly have problems of its own, and policymakers would need to implement it in a way that would minimize distortions to labor market conditions as much as possible. As we noted earlier, in the case of Europe, government involvement in the labor market often can have undesirable effects.

Perhaps an even more promising option would be to increase public investment in skill acquisition. As we have argued, the principal factor driving wage inequality is skill-biased technical change. Thus, the most direct and arguably most effective way to reduce such inequality would be to reduce the disparity in skills between workers.

What type of skills should we attempt to provide through public investment? The evidence seems increasingly clear that there is a relatively high level of return on investments in education early in life. As Pedro Carneiro and James Heckman write, "Skill and ability beget future skill and ability."[27]

---

[24] Sachs and Shatz (1996, 239).

[25] Kletzer and Litan (2001, 2).

[26] Kletzer and Litan outline such a proposal that would work as follows: Once displaced workers found new jobs, they would receive a subsidy to increase their current lower wage to a level more closely approximating their former higher wage. The wage subsidy would be available for only a limited period of time following reemployment and there would be an annual cap on payments. Ibid., p. 4.

[27] Carneiro and Heckman (2003).

Also, we might expect those investments to yield larger benefits if they are directed toward broadly generalizable skills. The ability to think critically, for instance, is crucial to analyzing and adapting to a number of situations. In contrast, the return on educational investments later in life, especially remedial or compensatory investments, tend to be smaller. This is true for at least two reasons. First, without a basic level of knowledge on which to build, it will be difficult for individuals to effectively acquire new skills. Second, by definition, older workers have less time to recoup the investment in education than younger workers.

While this may make perfect sense analytically, it still may be difficult to accept. Such reasoning implies that the people hurting the most now—those who have been displaced from their jobs—may also have the most trouble building their skills. What should we do to help those people? A good argument could be made that the government should act as a clearinghouse of information about job training programs, though we should be cautious about expanding such training programs given their limited success.[28] Similarly, we should be skeptical about providing greater financial assistance to displaced workers seeking education at community colleges and four-year institutions. There is already a wide array of educational subsidies in place, which have substantially reduced potential credit constraints for low- and middle-income people.[29]

Still, increased investment in skill acquisition is a policy option worth significant consideration. If done properly, it may be an effective tool in reducing wage inequality and could yield additional benefits to the economy, such as increasing workers' productivity.

6.   CONCLUSION

Wage inequality in the United States is large and has been growing during the past 30 years. The main cause, it appears, is skill-biased technical change. Those workers with high skill levels have experienced more rapid wage growth than less-skilled workers, some of whom have seen an actual decline in their real wages.

This development is cause for concern to many people who fear that a large share of the workforce no longer has a reasonable chance of achieving its goals, monetary and otherwise. Such concern is understandable. Indeed, the evidence suggests that, at present, less-skilled workers face formidable challenges in the labor market. As a society, we ought to consider investing more funds in skill development—especially early skill development—to ensure that as many people as possible have the basic tools necessary to succeed.

---

[28] See Kletzer (1998, 131–33).

[29] See Carneiro and Heckman (2002).

But we also need to remember that technical change is not necessarily skill-biased. There have been significant episodes where technical innovation appears to have been skill-replacing. From today's vantage point, it seems unlikely that we will return to such a world, but developments may lead us in that direction. Market economies, though highly efficient, often move in surprising and unpredictable directions.

Perhaps most important, we ought to focus not just on the distributional effects of technical change—important as they may be—but also on aggregate well-being. Technical change has fueled much of the economic growth of the past two centuries and raised living standards to levels once unimaginable.

J. Bradford DeLong has calculated that Real GDP per worker grew from roughly $13,700 in 1890 to about $65,000 in 2000. That's nearly a five-fold increase. And as DeLong has noted, that significantly understates our improvement in living standards. In 1890, people "could not buy modern entertainment or communications or transportation technologies." There were "no modern appliances, no modern buildings, no antibiotics, no air travel. An income of $13,700 today that must be spent exclusively on commodities already in use in the late nineteenth century is, for all of us, worth a lot less than $13,700."[30]

It's useful to consider the alternative to embracing technology. By 1400, China had invented many of the technologies that triggered the Industrial Revolution of the eighteenth century, such as moveable-type printing, the water-powered spinning machine, and the blast furnace. Tight state controls impeded the spread of those technologies, however, preventing them from being used to their full potential and inhibiting further innovation.[31] We are not suggesting that others are seriously proposing blocking the development and distribution of new technologies in the United States as China did centuries ago. But we do think it is important to understand how powerful a force technology can be for human well-being—and how counterproductive it can be to curtail its growth.

Despite the pain that technological change can cause workers in certain segments of the labor force, we should remember that, on net, technical change is good for the economy and good for people. We should not discourage or lament it.

---

[30] DeLong (2000, 14–15).
[31] See Landes (1998), especially pp. 51–59.

REFERENCES

Acemoglu, Daron. 1998. "Why Do New Technologies Complement Skills? Directed Technical Change and Wage Inequality." *Quarterly Journal of Economics* 113 (November): 1055–89.

——————. 2002. "Technical Change, Inequality, and the Labor Market." *Journal of Economic Literature* 40 (March): 7–72.

——————, Philippe Aghion, and Giovanni L. Violante. 2001. "Deunionization, Technical Change, and Inequality." *Carnegie-Rochester Conference Series on Public Policy* 55 (December): 229–64.

Autor, David H., Lawrence F. Katz, and Melissa S. Kearney. 2004. "Trends in U.S. Wage Inequality: Re-Assessing the Revisionists." Manuscript, MIT Department of Economics (August).

Autor, David H., Lawrence F. Katz, and Alan B. Krueger. 1998. "Computing Inequality: Have Computers Changed the Labor Market?" *Quarterly Journal of Economics* 113 (November): 1169–213.

Autor, David H., Frank Levy, and Richard J. Murnane. 2003. "The Skill Content of Recent Technological Change: An Empirical Exploration." *Quarterly Journal of Economics* 118 (November): 1279–333.

Borjas, George J., Richard B. Freeman, and Lawrence F. Katz. 1997. "How Much Do Immigration and Trade Affect Labor Market Outcomes?" *Brookings Papers on Economic Activity* (No. 1): 1–67.

Burtless, Gary T., Robert Z. Lawrence, Robert E. Litan, and Robert J. Shapiro. 1998. *Globaphobia: Confronting Fears about Open Trade*. Washington, D.C.: Brookings Institution Press.

Card, David, and John E. DiNardo. 2002. "Skill-Biased Technological Change and Rising Wage Inequality: Some Problems and Puzzles." NBER Working Paper No. 8769 (February).

Card, David, Thomas Lemieux, and W. Craig Riddell. 2003. "Unionization and Wage Inequality: A Comparative Study of the U.S., the U.K., and Canada." NBER Working Paper No. 9473 (January).

Carneiro, Pedro, and James J. Heckman. 2002. "The Evidence on Credit Constraints in Post-Secondary Schooling." NBER Working Paper No. 9055 (July).

——————. 2003. "Human Capital Policy." NBER Working Paper No. 9495 (February).

DeLong, J. Bradford. 2000. "Cornucopia: The Pace of Economic Growth in the Twentieth Century." NBER Working Paper No. 7602 (March).

Eckstein, Zvi, and Éva Nagypál. 2004. "The Evolution of U.S. Earnings Inequality: 1961–2002." Federal Reserve Bank of Minneapolis *Quarterly Review* 28 (December): 10–29.

Goldin, Claudia, and Lawrence F. Katz. 1999. "The Returns to Skill in the United States across the Twentieth Century." NBER Working Paper No. 7126 (May).

Goldin, Claudia, and Robert A. Margo. 1992. "The Great Compression: The Wage Structure in the United States at Mid-Century." *Quarterly Journal of Economics* 107 (February): 1–34.

Henig, Harry, and S. Herbert Unterberger. 1945. "Wage Control in Wartime and Transition." *American Economic Review* 35 (June): 319–36.

Hornstein, Andreas, Per Krusell, and Giovanni L. Violante. 2004. "The Effects of Technical Change on Labor Market Inequalities." Federal Reserve Bank of Richmond Working Paper No. 04-08 (December). Also: Forthcoming in *Handbook of Economic Growth*, ed. Philippe Aghion and Steven Durlauf. Amsterdam: North-Holland.

Humphrey, Thomas M. 2004. "Ricardo Versus Wicksell on Job Losses and Technological Change." Federal Reserve Bank of Richmond *Economic Quarterly* 90 (Fall): 5–24.

Juhn, Chinhui, Kevin M. Murphy, and Brooks Pierce. 1993. "Wage Inequality and the Rise in Returns to Skill." *Journal of Political Economy* 101 (June): 410–42.

Kletzer, Lori G. 1998. "Job Displacement." *Journal of Economic Perspectives* 12 (Winter): 115–36.

——————, and Robert E. Litan. 2001. "A Prescription to Relieve Worker Anxiety." Brookings Institution Policy Brief #73 (March).

Krugman, Paul. 1994. "Past and Prospective Causes of High Unemployment." Federal Reserve Bank of Kansas City *Economic Review* 79 (Fourth Quarter): 23–43.

Landes, David S. 1998. *The Wealth and Poverty of Nations: Why Some Are So Rich and Some So Poor.* New York: W.W. Norton.

Lee, David S. 1999. "Wage Inequality in the United States During the 1980s: Rising Dispersion or Falling Minimum Wage?" *Quarterly Journal of Economics* 114 (August): 977–1023.

Lerman, Robert I. 1997. "Reassessing Trends in U.S. Earnings Inequality." *Monthly Labor Review* 120 (December): 17–25.

Ljungqvist, Lars, and Thomas J. Sargent. 1998. "The European Unemployment Dilemma." *Journal of Political Economy* 106 (June): 514–50.

Mokyr, Joel. 1994. "Technological Change, 1700–1830." In *The Economic History of Britain Since 1700, Volume I: 1700–1860* (2nd ed.), ed. Roderick Floud and Donald McCloskey. New York: Cambridge University Press.

Sachs, Jeffrey D., and Howard J. Shatz. 1996. "U.S. Trade with Developing Countries and Wage Inequality." *American Economic Review* 86 (May): 234–39.

Schumpeter, Joseph. 1942. *Capitalism, Socialism, and Democracy*. New York: Harper & Brothers.

Simon, Julian L. 1999. *The Economic Consequences of Immigration* (2nd ed.). Ann Arbor, Mich.: University of Michigan Press.

Topel, Robert H. 1997. "Factor Proportions and Relative Wages: The Supply-Side Determinants of Wage Inequality." *Journal of Economic Perspectives* 11 (Spring): 55–74.

Violante, Giovanni L. 2002. "Technological Acceleration, Skill Transferability, and the Rise in Residual Inequality." *Quarterly Journal of Economics* 117 (February): 297–338.

Unemployment and Vacancy Fluctuations in the Matching Model: Inspecting the Mechanism

Andreas Hornstein, Per Krusell, and Giovanni L. Violante

T he state of the labor market, employment and unemployment, plays an important role in the deliberations of policymakers, the Federal Reserve Bank included. Over the last 30 years, economic theory has led to substantial progress in understanding the mechanics of business cycles. Much of this progress in macroeconomics has been associated with the use of calibrated dynamic equilibrium models for the quantitative analysis of aggregate fluctuations (Prescott [1986]). These advances have mainly proceeded within the Walrasian framework of frictionless markets. For the labor market, this means that while these theories contribute to our understanding of employment determination, they have nothing to say about unemployment.

Policymakers care about the behavior of unemployment for at least two reasons. First, even if one is mainly interested in the determination of employment, unemployment might represent a necessary transitional state if frictions impede the allocation of labor among production opportunities. Second, job loss and the associated unemployment spell represent a major source of income risk to individuals.

Over the past two decades, the search and matching framework has acquired the status of the standard theory of equilibrium unemployment.[1] This theory is built on the idea that trade in the labor market is costly and takes time. Frictions originating from imperfect information, heterogeneity

[1] For a textbook survey, see Pissarides (2000).

of firms and workers, and lack of coordination disrupt the ability to form
employment relationships. The quantity of idle inputs in the labor market
(unemployed workers and vacant jobs) is a measure of such disruption. In
its most basic representation, a labor market matching model focuses on the
interaction between unemployment and job creation. Higher productivity in-
creases the return to job creation and thereby increases the rate of job creation.
In turn, a higher rate of job creation makes it easier for unemployed workers
to find jobs and thereby reduces unemployment. This explains the observed
counter-cyclical (pro-cyclical) behavior of unemployment (job creation).

Shimer (2005) goes beyond investigating the qualitative features of the
basic matching model. He follows the research program on dynamic equilib-
rium models with Walrasian frictionless markets and explores whether or not
a calibrated matching model of the labor market is quantitatively consistent
with observed aggregate fluctuations. He surprisingly concludes that a reason-
ably calibrated matching model does not generate enough volatility in unem-
ployment and cannot explain the strong procyclicality of the job-finding rate.
In other words, the matching model stops short of reproducing the cyclical
behavior of its two central elements: unemployment and vacancies.[2]

In this article, we present the basic matching model, also known as the
Mortensen-Pissarides model, in detail and, building on Shimer (2005), we
explain the reasons for the quantitative problems of the model. Essentially,
given the way wages are determined in the (Nash-bargaining) model and the
way Nash bargaining is calibrated, wages respond strongly to changes in pro-
ductivity so that the incentive for firms to create jobs does not change very
much. We then discuss two possible ways of reconciling a matching model
with the data.

First, as argued by Hall (2005) and Shimer (2004), if wages are essentially
rigid, the model performs much better. We contend that rigid wages per se
are not sufficient; another necessary requirement is a very large labor share—
close to 100 percent of output. Moreover, we show that with rigid wages, the
model has implications for the labor share that seem too extreme: the labor
share becomes perfectly negatively correlated with—and as volatile as—labor
productivity whereas in the data this correlation is -0.5, and the variation of
the share is not nearly as large as that of productivity.

Second, as suggested by Hagedorn and Manovskii (2005), without aban-
doning Nash bargaining, a different calibration of some key parameters of the

---

[2] We should note that Andolfatto (1996) and Merz (1995) were the first to integrate the
matching approach to the labor market into an otherwise standard Walrasian model and to evaluate
this model quantitatively. Their work, however, was not so much focused on the model's ability to
match the behavior of unemployment, but on how the introduction of labor market frictions affects
the ability of the otherwise standard Walrasian model to explain movements in employment, hours
worked, and other non-labor-market variables. Andolfatto (1996), however, also pointed out the
model's inability to generate enough volatility in vacancies.

model also allows one to raise the volatility of unemployment and vacancies in the model. For this calibration to work, however, one again needs a very high wage share. This high share is obtained by "artificially" raising the outside option of the worker through generous unemployment benefits.[3] We tentatively conclude, as do Costain and Reiter (2003), that this parameterization has implausible implications for the impact of unemployment benefits on the equilibrium unemployment rate: a 15 percent rise in benefits would *double* the unemployment rate.

Why is a very large (very small) wage share (profit share) so important in order for the model to have a strong amplification mechanism for vacancies and unemployment? The model has a free-entry condition stating that vacancies are created until discounted profits equal the cost of entry. If profits are very small in equilibrium, a positive productivity shock induces a very large percentage increase in profits, and hence a large number of new vacancies must be created—through firm entry—thus lowering the rate of finding workers enough that entry remains an activity with zero net payoff.

We conclude that neither one of the solutions proposed is fully satisfactory, for two reasons. First, they both have first-order counterfactual implications. Second, they both assume a very large value for the labor share. It is hard to assess whether this value is plausible because there is no physical capital in the baseline matching model. We speculate that the addition of physical capital, besides providing a natural way of measuring the labor share of aggregate income, would allow the analysis of another important source of aggregate fluctuations, investment-specific shocks, which have proved successful in Walrasian models.[4]

The present article, which can be read both as an introduction to the matching model of unemployment and as a way of understanding the recent discussions of the model's quantitative implications, is organized as follows. We first quickly describe the data. Next, we describe in Section 2 the basic model without aggregate shocks. In Section 3, we define and solve for a stationary equilibrium: a steady state. In Section 4, we briefly discuss transition dynamics within the model without shocks. In Section 5, we derive the qualitative comparative statics for a one-time permanent change of the model's parameters. In Section 6, we present the alternative calibration strategies one could follow to parameterize the matching model, and in Section 7, we show how the quantitative comparative statics results differ according to the model

---

[3] To be precise, a large wage share is also sufficient for a strong amplification mechanism with rigid wages. With flexible wages, the large wage share must be achieved by making unemployment benefits high.

[4] See Greenwood, Hercowitz, and Krusell (2000) and Fisher (2003) for a quantitative account of the role of this type of shock in U.S. business cycles over the postwar period. Costain and Reiter (2003) illustrate quantitatively that productivity shocks affecting only new jobs improve the performance of the baseline model.

Table 1 Aggregate Statistics: 1951:1–2004:4

			HP Smoothing Parameter: 10^5				
	u	v	θ	λ_w	w	s	p
Standard Deviation	0.20	0.23	0.38	0.12	0.02	0.02	0.02
Autocorrelation	0.94	0.95	0.95	0.91	0.95	0.91	0.89
Correlation with p	−0.40	0.31	0.38	0.38	0.69	−0.35	1.00
			HP Smoothing Parameter: 1600				
	u	v	θ	λ_w	w	s	p
Standard Deviation	0.13	0.14	0.26	NA	0.01	0.01	0.01
Autocorrelation	0.87	0.90	0.89	NA	0.81	0.77	0.76
Correlation with p	−0.29	0.45	0.38	NA	0.72	−0.61	1.000

Notes: Data are quarterly, and u is the unemployment rate of the civilian population; v is the help-wanted advertising index; $\theta = v/u$ is labor market tightness; p is output per employee in the nonfarm business sector; s is the labor share constructed as the ratio of compensation of employees to output in the nonfarm sector; w is the wage computed as labor share times labor productivity, i.e., $w = s \cdot p$. The statistics for the job-finding rate, λ_w, are those reported in Shimer (2005) for an HP smoothing parameter of 10^5.

calibration. In Section 8, we introduce explicit stochastic aggregate shocks and discuss how the quantitative comparative statics results for one-time permanent shocks have to be modified to account for persistent but temporary shocks. Section 9 concludes the article.

1.   THE DATA

The focus of the analysis is on fluctuations at the business-cycle frequencies, and hence low-frequency movements in the data should be filtered out. For quarterly data, the standard practice (followed by Andolfatto and Merz) is to use a Hodrick-Prescott (HP) filter with a smoothing parameter set to 1600. Shimer (2005) chooses a much smoother trend component, corresponding to an HP smoothing parameter of 10^5.

Table 1 summarizes the key labor market facts around which this article is centered. We report statistics for the detrended log-levels of each series. When we remove a very smooth trend (smoothing parameter 10^5), we can summarize the data as follows:

- **Unemployment and Vacancies.** First, unemployment, u, and vacancies, v, are about 10 times more volatile than labor productivity, p. Market tightness, θ, defined as the ratio of vacancies to unemployment, is almost 20 times more volatile. Second, market tightness is

positively correlated with labor productivity. Both unemployment and vacancies show strong autocorrelation.

- **Job-Finding Rates.** The job-finding rate, λ_w, is six times more volatile than productivity and is pro-cyclical. It is also strongly autocorrelated.

- **Wages and Labor Share.** Wages and the labor share are roughly as volatile as labor productivity. The correlation between wages and labor productivity is high but significantly less than one, and the labor share is countercyclical.

Using a more volatile trend component (lower smoothing parameter) has almost no effect on the relative volatilities. For the vast majority of the variables, the percentage standard deviation is reduced roughly by one-third. Interestingly, the volatility is cut in half for wages and the labor share. Overall, the autocorrelation of the series is reduced, since some of the persistence is absorbed by the more variable HP trend. Finally, the correlation structure of the series with labor productivity is, in general, unchanged except for the labor share whose negative correlation almost doubles.

We conclude that the choice of smoothing parameter has no impact on the unemployment and vacancy statistics but does affect the labor share statistics somewhat.

2. THE MODEL

We now outline and discuss the basic Mortensen-Pissarides matching model with exogenous separations.[5] We choose a formulation in continuous time in order to simplify some of the derivations. It is useful to first describe the stationary economy (when aggregate productivity is constant over time) because that model is simple and yet very informative about how the model with random shocks behaves. Later, we will briefly discuss aggregate fluctuations with stochastic productivity shocks that are persistent but not permanent.[6]

Workers and Firms

There is a fixed number of workers in the economy; the model does not consider variations in the labor force or in the effort or amount of time worked by each

---

[5] The main reference is Pissarides (1985). Mortensen and Pissarides (1994) extend the model to endogenous separations. Pissarides (2000) contains an excellent survey of the matching models. See also Rogerson, Shimer, and Wright (2005) for a recent survey.

[6] The view that aggregate fluctuations in output and unemployment are due to fluctuations in productivity is not essential here. For the given environment, one can interpret productivity shocks as actually representing another source of fluctuations (such as "demand shocks," e.g., shocks to preferences).

worker. For example, think of workers as being uniformly distributed on the interval [0,1]—for any point on this interval, there is one worker—though there is no particular meaning to a worker's position on the interval.

Workers are all the same from the perspective of both their productivity and their preferences. Workers are infinitely lived and have linear utility over consumption of a homogeneous good, meaning that to the extent that there is uncertainty, workers are risk-neutral. There is constant (exponential) discounting at rate r. One can therefore think of a worker's expected present value of utility as simply the expected present value of income.[7]

Workers are either employed or unemployed. An employed worker earns wage income, w, but cannot search. Unemployed workers search for jobs. Let $b > 0$ denote the income equivalent of the utility flow that a worker obtains in the nonworking activity when unemployed, e.g., the monetary value of leisure plus unemployment benefits net of search costs.[8]

A firm is a job. The supply of firms (jobs) is potentially infinite. Every firm is equally productive at any point in time. Firms are risk-neutral and they discount future income at the same rate as do workers. Production requires one worker and one firm; firms can really be thought of as another type of labor input, such as an "entrepreneur." A firm-worker pair produces p units of the homogeneous output per unit of time. We assume that the value of production for a pair always exceeds the value of not working for a worker, i.e., that $p > b > 0$.[9] There is no cost for a firm to enter the labor market.

The Frictional Labor Market

In a "frictional" labor market, firms and workers do not meet instantaneously. In addition, firms that want to meet workers have to use resources to post a vacancy. In particular, a firm has to pay c units of output per unit of time it posts a "vacancy." Let the number of idle firms that have an open position be denoted $v(t)$, and let the number of unemployed workers be $u(t)$. Lack of coordination, partial information, and heterogeneity of vacancies and workers are all factors that make trading in the labor market costly.

We do not model these labor market frictions explicitly but use the concept of a matching function as a reduced form representation of the frictions.[10] This

---

[7] Alternatively, one could assume that workers are risk-averse but that they can obtain complete insurance against idiosyncratic income risk. In this case, it would also be optimal for workers to maximize the expected present value of income.

[8] Note that unemployment benefits do not serve an insurance role in this environment since workers are either risk-neutral or they already obtain complete insurance.

[9] This condition is necessary for ruling out a trivial equilibrium with zero employment: if $b > p$, no worker would be willing to work even if she could extract the entire value of the output produced from the firm.

[10] The concept of an "aggregate matching function" has been around for some time. In their survey of the literature on matching functions, Petrolongo and Pissarides (2001) include a short

formulation specifies that the rate at which new matches, m, are created is given by a time-invariant function, M, of the number of unemployed workers searching for a job and the number of vacant positions: $m = M(u, v)$. At this point, we will assume that M is (1) increasing and strictly concave in each argument separately and (2) constant returns to scale (CRS) in both arguments. Thus, matches are more likely when more workers and firms are searching, but holding constant the size of one of the searching groups, there are decreasing marginal returns in matching.

New matches are formed according to Poisson processes with arrival rates λ_w and λ_f. Given the rate at which new matches are formed, the rate at which an unemployed worker meets a firm is simply $\lambda_w(t) = m(t)/u(t)$, the total number of successful matches per worker searching. Similarly, the rate at which a vacant firm meets a worker is $\lambda_f(t) = m(t)/v(t)$. Since the matching function is CRS, the two meeting rates depend on labor market tightness, $\theta(t) = v(t)/u(t)$, only:

$$\lambda_w(t) = M[1, \theta(t)] \text{ and } \lambda_f(t) = M[1/\theta(t), 1]. \tag{1}$$

As the relative number of vacancies increases, the job-finding rate, λ_w, also increases, but the worker-finding rate, λ_f, decreases. We assume that once a firm and a worker have been matched, they remain matched until "separation" occurs. Separation occurs according to a Poisson process with exogenous arrival rate, σ.

If an unemployed worker meets vacant firms according to a Poisson process with arrival rate, λ_w, then the probability that the worker meets exactly one vacant firm during a time period, Δ, is $\lambda_w \Delta$ if the time period is sufficiently short. Furthermore, the probability that a worker meets two or more vacant firms during this time period is essentially zero.[11] Similarly, the probability that a vacant firm meets an unemployed worker is $\lambda_f \Delta$, and the probability that a matched firm-worker pair separates is $\sigma \Delta$. Thus, if we start out with $u(t)$ unemployed workers and $1 - u(t)$ employed workers at time t, after a short time period, Δ, the number of unemployed workers will be

$$u(t + \Delta) = \sigma \Delta [1 - u(t)] + [1 - \lambda_w(t) \Delta] u(t).$$

---

history of the concept. Lagos (2000) warns against the dangers of such a "reduced-form approach" to frictions when, for example, evaluating the effects of policies. The underlying reason is that policies may affect the search behavior of agents and change the shape of the aggregate matching function.

[11] Note that for a Poisson process, the rate λ at which the state changes need not be bounded above by one. Since we are interested in the limiting case when the time interval, Δ, becomes arbitrarily small, the probability of a state change, $\lambda \Delta$, will eventually be less than one for any fixed and finite λ.

Subtracting $u(t)$ from either side of this expression, dividing by Δ, and taking the limit when the length of the time period goes to zero, we obtain

$$\dot{u}(t) = \lim_{\Delta \to 0} \frac{u(t+\Delta) - u(t)}{\Delta} = \sigma[1 - u(t)] - \lambda_w(t)u(t). \qquad (2)$$

Here $\dot{u}(t)$ denotes the time derivative (change per unit of time) of $u(t)$: $\dot{u}(t) = \partial u(t)/\partial t$. This equation captures that the change in unemployment is the flow into unemployment (the number of employed workers times the rate at which they separate) minus the flow from unemployment (the number of unemployed workers times the rate at which they find a job).

The dynamic evolution of unemployment is one of the key concerns in this model. Notice, however, that the job-finding rate for workers, $\lambda_w(t)$, in equation (2) depends on vacancies through labor market tightness, $\theta(t)$. What determines vacancies, $v(t)$? In order to answer this question, we need to describe what determines profits for entering firms, which in turn requires us to discuss what wages workers receive.

With matching frictions, both workers and firms have some bargaining power since neither party can be replaced instantaneously, as is commonly assumed in competitive settings. There is a variety of theories that describe how bargaining allocates output between firms and workers under these circumstances. Below we will determine wages according to the widely used Nash-bargaining solution. For simplicity, from now on we will mainly consider steady states, situations in which all aggregate variables are stationary over time. Thus, $u(t)$, $v(t)$, $\lambda_w(t)$, and $\lambda_f(t)$ are all constant, even though individual workers and firms face uncertainty in their particular experiences.

3. STATIONARY EQUILIBRIUM

Values

Denote the net present value of a matched firm J (which in general would depend on time but in a steady state does not). Given output, p, and the wage, w, paid to its worker, J must satisfy

$$rJ = p - w - \sigma(J - V), \qquad (3)$$

where V is the value of the firm when unmatched. This equation is written in flow form and is interpreted as follows: the flow return of being matched—the capital value of being matched times the rate of return on that value—equals the flow profits minus the expected capital loss resulting from match separation—the rate at which the firm is separated, σ, times the latter capital loss equals $J - V$.[12]

---

[12] This equation is written in flow form but can be derived from a discrete-time formulation analogous to the derivation of equation (2). Suppose that the value of being vacant is constant

Similarly, the value of a vacant firm satisfies

$$rV = -c + \lambda_f (J - V). \qquad (4)$$

Here, there is a flow loss due to the vacancy posting cost and an expected capital gain from the chance of meeting a worker.

Turning to the net present value of a matched worker, W, and an unemployed worker, U, we similarly have

$$rW = w - \sigma (W - U), \text{ and} \qquad (5)$$
$$rU = b + \lambda_w (W - U). \qquad (6)$$

The flow return from not working, b, could be a monetary unemployment benefit collected from the government, a monetary benefit from working in an informal market activity, or the monetary equivalent of not working in any market activity (the value of being at home). We will discuss the role and interpretation of b more extensively below, because it turns out that it matters how one thinks of this parameter.

Wage Determination

The values of (un)matched workers and firms depend on the wages—yet to be determined—paid in a match. Obviously, for a match to be maintained it must be beneficial for both the worker, $W - U \geq 0$, and the firm, $J - V \geq 0$. We define the total surplus of a match, $S \equiv (J - V) + (W - U)$, as the sum of the gain of the firm and worker being in a match relative to not being in a match. We assume that the wage is set such that the total match surplus is shared between the worker and firm according to the Nash-bargaining solution with share parameter β:[13]

$$W - U = \beta S \text{ and } J - V = (1 - \beta)S. \qquad (7)$$

---

over time from the perspective of a matched firm and that we are looking at one period being of length Δ. During this period, there is production, and wages are paid, the net amount being $(p - w)\Delta$ since p and w are measured per unit of time. At the end of the period, the match separates with probability $\sigma\Delta$ and remains intact with probability $1 - \sigma\Delta$. So it must be that $J(t) = (p - w)\Delta + (1 - \sigma\Delta)e^{-r\Delta} J(t + \Delta) + \sigma\Delta e^{-r\Delta} V$. Here, $e^{-r\Delta} \equiv \delta(\Delta)$ is a discount factor; it gives a *percentage* decline in utility as a function of the length of time, $-(d\delta(\Delta)/d\Delta)/\delta(\Delta)$, which is constant and equal to r. Subtract $J(t + \Delta)e^{-r\Delta}$ on both sides and divide by Δ. That delivers $\frac{J(t) - J(t+\Delta)}{\Delta} + \frac{(1 - e^{-r\Delta})}{\Delta} J(t + \Delta) = p - w - \sigma e^{-r\Delta}(J(t + \Delta) - V)$. Take limits as $\Delta \to 0$. Then the left-hand side becomes $\dot{J}(t) + rJ(t)$, the second term coming from an application of l'Hôpital's rule and the value being a continuous function of time. The right-hand side gives $p - w - \sigma(J(t) - V)$. In a steady state, $J(t)$ is constant and equal to J, satisfying the equation in the text.

[13] The Nash-bargaining solution does not describe the outcome of an explicit bargaining process; rather, it describes the unique outcome among the set of all bargaining processes whose outcomes satisfy certain axioms (Nash [1950]). Also, one can derive the Nash-bargaining solution as the outcome of a bargaining process where participants make alternating offers until they reach agreement. For a survey of the bargaining problem, see Osborne and Rubinstein (1990).

Summing the value equations for matched pairs and subtracting the values of unmatched firms and workers, using the Nash-bargaining rule, we therefore obtain

$$rS = p - \sigma S + c - \lambda_f(1 - \beta)S - b - \lambda_w \beta S, \qquad (8)$$

which implies that

$$S = \frac{p + c - b}{r + \sigma + (1 - \beta)\lambda_f + \beta\lambda_w}. \qquad (9)$$

That is, we can express the surplus as a function of the primitives and the matching rates, which are endogenous and will be determined by the free entry of firms as shown below. We see that the surplus from being in a match is

- decreasing in the interest rate (a higher interest rate reduces the present value of remaining in the match),

- decreasing in the separation rate (a higher separation rate lowers the expected value of remaining together),

- decreasing in the bargaining share of workers times the rate at which they meet vacant firms (the higher the chance that unemployed workers meet vacant firms and the higher the share that workers receive in that case, the less valuable it is to be matched now), and

- decreasing in the bargaining share of firms times the rate at which vacant firms meet unemployed workers (the higher the chance that vacant firms meet unemployed workers and the higher the share that firms receive in that case, the less valuable it is to be matched now).

To derive a useful expression for the wage, subtract rV from the value equation for matched firms, (3), and use the Nash-bargaining rule to obtain

$$r(1 - \beta)S = p - w - \sigma(1 - \beta)S - rV. \qquad (10)$$

Also, notice that given the surplus sharing rule, (7), and the expressions for the vacancy and unemployment values, (4) and (6), the surplus in (8) can be written as

$$rS = p - \sigma S - rV - rU. \qquad (11)$$

Now multiply equation (11) by $1 - \beta$, subtract it from equation (10), and solve for the wage:

$$w = \beta(p - rV) + (1 - \beta)rU. \qquad (12)$$

Thus, the wage is a weighted average of productivity minus the flow value of a vacancy and the flow value of unemployment with the weights being β and $1 - \beta$, respectively. Intuitively, one can understand this equation as follows:

$w - rU$, the flow advantage of being matched for a worker, is just its share, β, of the overall advantage of being matched for the worker and the firm together, $\beta(p - rV - rU)$.

Firm Entry

There is an infinite supply of firms that can post vacancies, and entry is costless. Therefore, in an equilibrium with a finite number of firms posting vacancies, the value of a posted vacancy is zero:

$$V = 0. \tag{13}$$

If $V < 0$, no firm would enter, and if $V > 0$, an infinite number of firms would enter. This means that the number of vacancies, $v(t)$, adjusts at each point in time so that there are zero profits from entering, given the matching rate with workers, λ_f, which depends on $u(t)$ and on $v(t)$.

The free-entry condition (13), together with the definition of the vacancy value (4) and the surplus sharing rule (7) then determine the surplus value:

$$S = \frac{c}{(1 - \beta)\lambda_f}. \tag{14}$$

Moreover, we can use the free-entry condition to simplify the expression for the surplus in (9); the surplus can now be expressed as

$$S = \frac{p - b}{r + \sigma + \beta\lambda_w}. \tag{15}$$

These two expressions for the surplus can be combined to write

$$\frac{p - b}{r + \sigma + \beta\lambda_w} = \frac{c}{(1 - \beta)\lambda_f}. \tag{16}$$

This is an equation in one unknown, labor market tightness (θ), since both meeting rates (λ_w and λ_f) depend only on the number of vacancies relative to the unemployment rate (see equation (1)).

We also see that free entry implies that the wage expression (12) simplifies to

$$w = \beta p + (1 - \beta)rU. \tag{17}$$

Equilibrium Unemployment

In a steady state, $\dot{u}(t) = 0$, so the evolution for unemployment as given by equation (2) becomes

$$\sigma(1 - u) = \lambda_w u. \tag{18}$$

Thus, in a steady state, the flow into unemployment—the separation rate in existing matches times the number of matches—must equal the flow out of unemployment—the job-finding rate times the number of unemployed.

The steady state expression for unemployment can, on the one hand, be used to express unemployment as a simple function of the separation rate and the job-finding rate. On the other hand, it can be used to write the job-finding rate in terms of the unemployment rate and the separation rate. If we know, for example, that the unemployment rate is 10 percent and that the monthly separation rate is 5 percent, then the chance of finding a job within a month must be $\sigma \frac{1-u}{u} = 0.05 \cdot \frac{0.9}{0.1} = 0.45$; that is, just under one-half.

Solving the Model

Solving the model is now straightforward. We have derived (16) and (18) in two unknowns, θ and u. Furthermore we can solve the two equations sequentially. First, from (1) it follows that λ_w (λ_f) is increasing (decreasing) in θ. This, in turn, implies that the left-hand side (LHS) of (16) is decreasing in θ and that the right-hand side (RHS) is increasing in θ. Thus, if a solution, θ, to (16) exists, it is unique. Second, conditional on θ, we can solve (18) for the equilibrium unemployment rate.

One can show that a solution to (16) exists if we assume that the matching function satisfies the Inada conditions.[14] We assume a particular functional form for the matching function that meets these conditions and that is the most common one in the literature, the Cobb-Douglas (CD) matching function,

$$M(u, v) = A u^\alpha v^{1-\alpha}. \tag{19}$$

The CD matching function has convenient properties in terms of how the matching rates change with changes in labor market tightness,

$$\lambda_w = A \theta^{1-\alpha} \text{ and } \lambda_f = A \theta^{-\alpha}. \tag{20}$$

Independent of the level of unemployment, if the labor market tightness increases by 1 percent, the rate at which a worker (firm) finds a firm (worker) goes up (down) by $1 - \alpha$ (α) percent. [15]

Using the CD matching function, our equilibrium condition, (16), becomes

$$\frac{p - b}{r + \sigma + \beta A \theta^{1-\alpha}} = \frac{c}{(1 - \beta) A \theta^{-\alpha}}. \tag{21}$$

For $\theta = 0$, the LHS of (21) is finite and positive, and the RHS is zero. As θ becomes arbitrarily large, the LHS converges to zero and the RHS becomes arbitrarily large. Thus there exists a positive θ that solves (21). The unem-

---

[14] Let $f(\theta) = M(\theta, 1)$. Then the Inada conditions are $f(0) = 0$, $f(\infty) = \infty$, and $f'(0) = \infty$.

[15] Shimer (2005) argues that the constant elasticity CD matching function describes the data for the U.S. labor market well. See also Section 7 on calibration.

ployment rate then can be solved for in a second step, using (18), as

$$u = \frac{\sigma}{\sigma + A\theta^{1-\alpha}}. \tag{22}$$

We obtain the wage by using the definitions of the matching rates, (20), and substituting the expressions for rU and the value of S, (6), and (15) in wage equation (17):

$$w = \beta p + (1 - \beta)\left(b + \frac{\beta c}{1 - \beta}\theta\right) = \beta(p + c\theta) + (1 - \beta)b. \tag{23}$$

A Digression: The Frictionless Model

We now show that as search frictions become small, the equilibrium of the economy with matching frictions converges to the equilibrium of the corresponding economy without matching frictions. Search frictions can become small either because the cost of searching for vacant firms, c, becomes small or because the efficiency of the matching process, A, improves.

The frictionless economy is identical to the one outlined so far, except that matching between vacant firms and unemployed workers is instantaneous and costless. The resource allocation problem in the frictionless economy, which can be studied from the perspective of a benevolent social planner, is trivial. There will always be the same number of firms as workers operating because there is no cost in creating vacancies, and the matching process is instantaneous. Leaving workers idle would therefore be inefficient since $p > b$. There are no vacancies since matching is instantaneous. There is a competitive equilibrium that supports this allocation given some wage rates, $w(t)$, specified at all points in time. It is clear that for these wages, $w(t)$ must equal p for all t because workers are in short supply, and firms are not. That is, firm entry bids down profits to zero, and workers obtain the entire output.

Now suppose that the vacancy-posting costs become arbitrarily small: $c \to 0$. Then for any finite θ, the LHS of (21) is strictly positive, but the RHS converges to zero. Therefore, it must be that $\theta \to \infty$. To find the wage, some care must be taken, since the wage expression contains $c\theta$, i.e., $0 \cdot \infty$. Since workers meet firms at an ever-increasing rate, $\lambda_w \to \infty$, the unemployment rate becomes arbitrarily small, $u \to 0$, and from equation (9) it follows that the surplus from being in a match becomes arbitrarily small: $S \to 0$. Then simply inspect (10), which implies that $w \to p$, as expected: workers obtain the whole production value.

The same kind of result is obtained if the matching efficiency becomes arbitrarily large, $A \to \infty$. Now, however, there will be no vacancies, and θ will remain finite. To see this formally, multiply (21) with $A\theta^{-\alpha}$, divide the numerator and denominator of the LHS by A, and take the limit as $A \to \infty$:

$$\frac{p - b}{\beta\theta_\infty} = \frac{c}{1 - \beta}.$$

Since $\theta_\infty = \lim_{A \to \infty} \theta(A)$ is finite, the limits of both λ_f and λ_w are infinite. Thus from equation (9) it follows that the limit of the surplus is zero; from (22) it follows that the limit of the unemployment rate is zero; and from (10) it follows that the limit of the wage again must equal p. Since θ_∞ is positive and finite, v_∞ must equal 0 since u_∞ equals 0. There is no unemployment, and there are no vacancies.

4.   TRANSITION DYNAMICS

So far, we have discussed how the key endogenous variables—unemployment, vacancies, job-finding rates, and wages—are determined in steady state. But how does the economy behave out of a steady state? To answer this question, one needs to find out what the economy's state variables are. A state variable is a variable that is predetermined at time t and that matters to outcomes. Here, unemployment is clearly a state variable, because it is a variable that moves slowly over time according to (2). In fact, it is the only state variable. No other variable is predetermined. This means that, in general, allocations at t depend on $u(t)$ but not on anything else.

  So what is a dynamic equilibrium path of the economy if it starts with an arbitrary $u(0)$ at time zero? It turns out that the equilibrium is very easy to characterize. All variables except $u(t)$ and $v(t)$ will be constant over time from the very beginning.[16] To show that this is indeed an equilibrium, simply assume that θ is constant from the beginning of time and equal to its steady state value and then verify that all equilibrium conditions are satisfied. Since θ is constant, all job-finding rates—$\lambda_w(t)$ and $\lambda_f(t)$—will be constant and equal to their steady state values because they depend on θ and on nothing else. Since the λs are the only determinants of the values J, V, W, and U, the solution for the values will be the same as the steady state solution. It then also follows that the wage must be the steady state wage. To find $u(t)$ and $v(t)$, we conclude that $u(t)$ will simply follow

$$\dot{u}(t) = \sigma [1 - u(t)] - \lambda_w u(t), \tag{24}$$

which differs from (2) only in that λ_w is now constant. Once we have solved for $u(t)$, we can find the path for $v(t)$ residually from $v(t) = \theta u(t)$. Moreover, note that if $u(0)$ is above the steady state, u, the RHS of equation (24) is negative, which means that $\dot{u}(0)$ is negative. Unemployment falls, and as long as it is still above u, it continues falling until it reaches (converges to) u. Similarly, if it starts below u, it rises monotonically over time toward u.[17]

---

[16] Pissarides (1985; 2000, Chapter 1) shows that this is the unique equilibrium path.

[17] Formally, the solution for $u(t)$ is the solution to the linear differential equation (24): $u(t) = u + e^{-(\sigma + \lambda_w)t}(u(0) - u)$, where $u = \frac{\sigma}{\sigma + \lambda_w}$.

The fundamental insight here is that there are no frictions involved in firm entry, but there are frictions in movement of workers in and out of jobs.[18] Therefore, $u(t)$ is restricted to follow a differential equation which is "slow-moving," whereas $v(t)$ does not have to satisfy such an equation. It can jump instantaneously to whatever is has to be so that θ is equal to its steady state value from the beginning of time.

5. COMPARATIVE STATICS

We now analyze how different parameters influence the endogenous variables. In particular, how does unemployment respond to changes in productivity? Here, we emphasize that these are steady state comparisons. We find the long-run effect of the permanent change in the parameter. For most variables—all except $u(t)$ and $v(t)$—the impact of a permanent change in the parameter is instantaneous because θ immediately moves to its new, long-run value (see the discussion in the previous section). Of course, in the section below where some of the primitives are stochastic, their changes need not be permanent, and slightly different results apply.

For example, if we are looking at a 1 percent permanent increase in productivity, p, the comparative statics analysis in this section will correctly describe the effect on θ both in the long and in the short run, whereas the effect on unemployment recorded here only pertains to how it will change in the long run. The short-run effect on unemployment of a permanent change in a parameter is straightforward to derive, nevertheless: It simply involves tracing out the new dynamics implied by the linear differential equation (24) evaluated at the new permanent value for λ_w (which instantaneously adopts its new value because θ does). In particular, one sees from the differential equation that an increase in θ will increase λ_w and thus increase the speed of adjustment to the new steady state rate of unemployment.

We are mainly interested in how the economy responds to changes in p, but we will also record the responses to b, σ, and c. We compute elasticities, i.e., we use percentage changes and ask by what percent θ and u will change when p, b, σ, and, c change by 1 percent. We derive the relevant expressions by employing standard comparative statics differentiation of (21) and (22). Using $\hat{x}$ to denote $d\log(x) = dx/x$, it is straightforward to derive

---

[18] The speed of movements from unemployment into employment is regulated by the hiring rate, λ_w, which, in turn, depends on the endogenous market tightness, θ. Separations instead are exogenous, and, hence, the speed of movements from employment to unemployment is simply determined by the parameter, σ.

$$\hat{\theta} = \frac{r + \sigma + \beta\lambda_w}{\alpha\,(r+\sigma) + \beta\lambda_w} \left[\frac{p}{p-b}\hat{p} - \frac{b}{p-b}\hat{b} - \frac{\sigma}{r+\sigma+\beta\lambda_w}\hat{\sigma} - \hat{c} \right], \text{ and}$$

(25)

$$\hat{u} = (1-u)\left[\hat{\sigma} - (1-\alpha)\,\hat{\theta}\right].$$

(26)

The Effect of an Increase in Productivity

From equation (25), we see that an increase in p of 1 percent leads to more than a 1 percent increase in θ since $\alpha < 1$, and $p > b > 0$. Intuitively, p increases the value of matches, and given that firms capture some of the benefits of this increase in value, there will be an increase in the number of firms per worker seeking to match. The larger the fraction of the surplus going to the firm (β small), the more vacancies and market tightness will respond to a change in labor productivity. We also see that to the extent that b is close to p, the effect can be large, since $p/(p-b)$ can be arbitrarily large. Why is this effect larger the closer b is to p? When $(p-b) \simeq 0$, the profit from creating vacancies is small, and $\theta \simeq 0$. Hence, even a small change in p induces very large changes in firms' profits and market tightness, θ, in percentage terms, through the free-entry condition (21).

　　Because the job-finding rate, λ_w, equals $A\theta^{1-\alpha}$, we obtain that $\hat{\lambda}_w = (1-\alpha)\hat{\theta}$, so the effect of p on θ is higher than that on job-finding rates by a constant factor, $1/(1-\alpha)$. If we look at the effect on unemployment, note from (26) that a 1 percent increase in θ lowers unemployment by $(1-u)(1-\alpha)$ percent.

The Effects of Changing b, σ, and c

Changes in income when unemployed, b, have a very similar effect to productivity changes, p, but with an opposite sign. Increasing b, in particular, lowers θ significantly if b is near p, but it has very little effect on θ if b is close to zero. An increase in the match separation rate, σ, decreases labor market tightness. More frequent separations reduce the expected profits from creating a vacancy, and, thus, θ falls. The effects on labor market tightness of higher vacancy-posting costs, c, are negative as well. A 1 percent increase in the vacancy cost lowers the labor market tightness (by less than 1 percent) because it requires firms' finding rates to go up in order to preserve zero profits, and, hence, there must be fewer vacant firms relative to unemployed workers. There is less than a one-for-one decrease because the surplus, once matched, increases as well, as is clear from equations (14) and (15).

The effects on the job-finding rate of all the above changes in primitives are all one minus α times the effect on $\hat{\theta}$. Similarly, the effects on unemployment are $-(1-u)(1-\alpha)$ times those on $\hat{\theta}$, with the exception of a change in σ because from (22), the total effect on unemployment of a rise in σ by 1 percent is twofold. The first effect is an indirect decrease through the impact on θ (a higher σ leads to a higher θ), which lowers unemployment. The second effect is a direct increase of unemployment due to the higher rate at which matches separate. The total effect cannot be signed without more detailed assumptions; for example, if $\alpha \geq 1/2$, the net effect is to increase unemployment.

An Additional Friction: Rigid Wages

In the model just described, productivity changes arguably have such a small impact on labor market tightness and unemployment that they cannot account for the observed fluctuations in the data. Hall (2005) and Shimer (2004) suggest that one way to address this shortcoming is to change the wage-setting assumption. We now describe a very simple model that captures this idea.

The values for workers and entrepreneurs continue to be defined by equations (3), (4), (5), and (6). Now, assume that wages are fixed at some exogenous level, $\bar{w}$, such that the implied capital values for entrepreneurs and workers satisfy $J > 0$ and $W > U$. Hall (2005) justifies this assumption on wage determination as a possible sustainable outcome of a bargaining game. The new equilibrium zero-profit condition from a vacancy creation is

$$\frac{p - \bar{w}}{r + \sigma} = \frac{c}{\lambda_f} = \frac{c}{A\theta^{-\alpha}}. \tag{27}$$

It follows that the impact of a change in labor productivity on market tightness is given by

$$\hat{\theta} = \frac{p}{\alpha\,(p - \bar{w})}\,\hat{p}. \tag{28}$$

Comparing this last expression to that in equation (25), we see that the rigid-wage model gives a stronger response. In particular, independent of b, if the average wage, $\bar{w}$, is large as a fraction of output (i.e., if the labor share is large), then market tightness will be very sensitive to small changes in productivity.

The effect on unemployment, given the changes in θ, is the same whether or not wages are rigid, as given by equation (26). Finally, a comparison of equations (21) and (27), reveals that by choosing a value for the worker's bargaining power, β, close to zero in the model with Nash bargaining, one achieves essentially rigid wages, since w is then almost the same as b.

Table 2 Parameters and Steady States for Calibrations

	Common across Calibrations	
	$r = 0.012, \alpha = 0.72, \ p = 1,$ $A = 1.35, \lambda = 1.35, \theta = 1, u = 0.07$	
	Specific to Calibrations	
	Shimer	Hagedorn & Manovskii	Hall
β	0.72	0.05	NA
b	0.40	0.95	0.40
w/p	0.98	0.97	0.98
b/w	0.41	0.98	0.41
η_{wp}	1.00	0.50	0.00

6.   CALIBRATION

In the previous section on comparative statics we demonstrated how steady states change when primitives change. In particular, we have analyzed qualitatively how a permanent productivity change affects labor market tightness (recall that the effect is the same in the short as well as in the long run) and how it influences unemployment in the long run. However, what are the magnitudes of these effects? In order to answer this question we need to assign values to the parameters, and we will do this using "calibration." We will, to the extent possible, select parameter values based on long-run or microeconomic data. Hence, we will not necessarily select those parameters that give the best fit for the time series of vacancies and unemployment, since we restrict the parameters to match other facts.

The parameters of the model are seven: $\beta, b, p, \sigma, c, A,$ and α. The steady state equations that one can use for the calibration are three: (21), (22), and (23). Some aspects of the calibration are relatively uncontroversial, but as we will see below, some other aspects are not. Therefore, we organize our discussion in two parts. We first describe how to assign values to the subset of parameters that allows relatively little choice. We then discuss the remaining parameters and show how, depending on what data one uses to calibrate these, different parameter selections may be reasonable. We also explain why this is a crucial issue—the effect of productivity changes for vacancies, and unemployment may differ greatly across calibrations. We summarize the different calibration procedures in Table 2.

Basic Calibration...

In this section, we follow the calibration in Shimer (2005). We think of a unit of time as representing one quarter. Therefore, it is natural to select $r = 0.012$,

given that the annual real interest rates have been around 5 percent. We choose the separation rate, $\sigma = 0.10$, based on the observation that jobs last about two and a half years on average.[19]

Job-finding rates in the data are estimated by Shimer to be 0.45 per month. Thus, a target for λ_w of 1.35 per quarter seems reasonable. Notice from equations (25), (26), and (28) that the response of labor market tightness and the unemployment rate to changes in productivity and other parameters does not depend on the worker-finding rate, λ_f. We therefore follow Shimer and simply normalize labor market tightness, $\theta = 1$, so that the worker-finding rate is equal to the job-finding rate.[20]

Next, consider the elasticity of the matching function: what should α be? Shimer plots the logarithm of job-finding rates against $\log(v/u)$ and observes something close to a straight line with a slope coefficient of about 0.28, which the theory's formulation, $\lambda_w = A\theta^{1-\alpha}$, says it should be. Therefore, we set $\alpha = 0.72$. Since we have set θ equal to one and λ_w equal to $A\theta^{1-\alpha}$, it follows that $A = 1.35$. From the condition determining steady state unemployment, (22), we now obtain that $0.1(1.35 - u) = u$, so that u is 6.9 percent, which is roughly consistent with the data. Notice also that the system of equilibrium conditions is homogeneous of degree one in c, p, and b. Therefore, we normalize $p = 1$ in steady state.

It remains to select c, b, and β. We have two equations left: the wage equation, (23), and the free-entry equilibrium condition, (21), which is the one that solves for θ in terms of primitives. We can think of this latter equation as residually determining c once b and β have been selected. Two more aspects of the data therefore need to be used in order to pin down b and β.

...but what are b and β?

We now turn to the more contentious part of the calibration.

Completing Shimer's Calibration

It is common to regard b as being the monetary compensation for the un-employed. The OECD (1996) computes average "replacement rates" across countries, i.e., the ratio of benefits to average wages, and concludes that, whereas typical European replacement rates can be up to 0.70, replacement

---

[19] For a Poisson process with arrival rate σ, the average time to the arrival of the state change is $1/\sigma$. Thus, the average time from forming a match to separation is $1/\sigma = 10$ quarters.

[20] Alternatively, we could have followed Hall (2005) and set the monthly worker-finding rate to one so that $\lambda_f = 3$, implying that $\theta = 1/3$. The value chosen for θ does not influence our results.

rates are at most 0.20 in the United States.[21] Shimer (2005) sets b equal to 0.4, which is even beyond this upper bound for the replacement rate since it turns out that the wage is close to one in his calibration. One reason why b should be higher than 0.2 is that it also includes the value of leisure associated with unemployment. We will discuss some alternative ways to calibrate b below.

Regarding β, it is common to appeal to the Hosios condition for an efficient search.[22] This condition says that in an economy like the present one, firm entry is socially efficient when the surplus sharing parameter, β, is equal to the elasticity parameter of the matching function, α. Thus, Shimer (2005) assumes that $\beta = \alpha$. This is one possible choice, though it is not clear why one should necessarily regard the real-world search outcome as efficient. In conclusion, if $\beta = 0.72$ and $b = 0.4$, from the free-entry condition we obtain $c = 0.324$, and the calibration in Shimer (2005) is completed. Note that Shimer does not use the wage equation in his calibration.

Alternative: Use the Wage Equation

Let us now look at an alternative way of calibrating the model that exploits the wage equation. Hagedorn and Manovskii (2005) point to two observations that arguably can be used to replace those used by Shimer to calibrate b and β.

First, they argue that one can look at the size of profits in the data. Referring to empirical studies, Hagedorn and Manovskii argue that the profit share, which they identify as $(p - w)/p$ in the model, is about 0.03.[23] That is, this calibration strategy is equivalent to selecting a wage share a few percentage points below one. Second, they argue that one can look at how much wages respond to productivity. Using microeconomic data, Hagedorn and Manovskii conclude that a 1 percent productivity increase raises wages by half a per-

---

[21] In the United States, unemployment insurance replaces around 60 percent of past earnings, but in the data, unemployed workers earn much less than the average wage.

[22] See Hosios (1990). Free entry of firms involves an externality since individual vacant firms do not take into account that variations in the vacancy rate affect the rate at which they meet unemployed workers and the rate at which unemployed workers meet them.

[23] A pure aggregate profit measure should probably take the cost of vacancies into account, and, as such, it should be computed somewhat differently:

$$((1 - u)(p - w) - vc)/(p(1 - u)) = 1 - (w/p) - \theta(c/p)(u/(1 - u)).$$

If this expression equals 0.03, one obtains a smaller wage share, but since c must be less than one for zero profits to be feasible, w/p cannot be below $0.97 - 1 \cdot 1 \cdot 0.05/0.95 \sim 0.92$. Thus, both computations lead to a wage share close to one.

cent.[24] We now show how one can use these two observations to determine b and β.

The wage share. The wage income share in the model is obtained by dividing the wage equation (23) by productivity:

$$\frac{w}{p} = \beta \left(1 + \frac{c\theta}{p} \right) + (1 - \beta) \frac{b}{p}. \tag{29}$$

Rearranging the equilibrium condition (21) yields

$$\frac{c\theta}{p} = \frac{(1 - \beta) \lambda_w}{r + \sigma + \beta \lambda_w} \left(1 - \frac{b}{p} \right). \tag{30}$$

It is informative to calculate the wage share implied by Shimer's calculations. Now, given Shimer's preferred parameter values,

$$\frac{c\theta}{p} \approx \frac{1 - \beta}{\beta} \left(1 - \frac{b}{p} \right)$$

since $(r + \sigma)$ is small relative to λ_w. Therefore, with this expression inserted into (29), we conclude that

$$\frac{w}{p} \approx 1,$$

meaning that calibration of the wage share to 0.97 will not *by itself* be a large departure from Shimer's parameterization. Indeed, Shimer obtains a wage share of $w/p = 0.973$.

However, there are several different choices of the pair $(b/p, \beta)$ that can achieve this value of the labor share. To see this, combine equations (29) and (30) by eliminating $c\theta/p$:

$$\frac{w}{p} = \frac{(r + \sigma) [\beta + (1 - \beta) b/p] + \beta \lambda_w}{(r + \sigma) + \beta \lambda_w}. \tag{31}$$

Shimer chooses a relatively large value of β, which makes the wage share in (31) close to one without imposing constraints on b/p. Alternatively, β can be set close to zero, in which case a value for b/p needs to be around one. Recall that with b close to p, the dynamic properties of the model change dramatically. The model has a much stronger amplification mechanism, but how can one justify this choice of β?

The wage elasticity with respect to productivity. We differentiate (31) in order to derive a relation between $\eta_{\theta p} \equiv d \log \theta / d \log p$, the percentage change in θ in response to a 1 percent increase in p, and $\eta_{wp} \equiv d \log w / d \log p$ (the

---

[24] When we regress the cyclical component of wages on labor productivity (see Table 1 for a description of the data), we obtain an elasticity of 0.57 with the low smoothing parameter and 0.72 with Shimer's smoothing parameter. The first number is higher than, but not too distant from, Hagerdorn and Manovskii's preferred estimate of 0.5. In particular, it is not statistically different from 0.5.

corresponding measure for how wages respond to productivity). We obtain

$$\eta_{wp} = \frac{\beta \left[1 + (c\theta/p) \, \eta_{\theta p} \right]}{w/p}. \tag{32}$$

When $r + \sigma$ is small relative to $\beta\lambda_w$, as in Shimer's calibration, the elasticity of labor market tightness with respect to productivity satisfies

$$\eta_{\theta p} \approx \frac{1}{1 - b/p},$$

demonstrating that the wage elasticity must be

$$\eta_{wp} \approx \frac{1}{w/p}$$

(it must be close to one if the labor share is near one). That is, Shimer's calibration generates a one-for-one wage increase in response to productivity, measured in percentage terms, which is twice as large as the estimates cited by Hagedorn and Manovskii.

To obtain such a low elasticity, one needs to decrease β, so that $r + \sigma$ is no longer small relative to $\beta\lambda_w$, and this is how Hagedorn and Manovskii accomplish the task. A combination of (32) and the exact expression for $\eta_{\theta p}$ from (25) allows us, after some simplifications, to solve for η_{wp} as

$$\eta_{wp} = \left(\frac{\beta}{w/p} \right) \left[\frac{\alpha(r + \sigma) + \lambda_w}{\alpha(r + \sigma) + \beta\lambda_w} \right]. \tag{33}$$

It is now easy to see that using the baseline (uncontroversial) calibration together with $w/p \approx 1$ and $\beta = 0.13$ takes us to a number for η_{wp} that is closer to one-half.[25] Notice also that when β is close to zero, the approximation that $\eta_{\theta p} \approx p/(p - b)$ is no longer so good; rather, $\eta_{\theta p}$ is significantly higher than $p/(p - b)$, thus further strengthening the amplification of shocks in the model.

Put differently, if we restrict the model so that it generates a weaker response of wages to productivity, then expression (33) tells us that β has to be significantly smaller. And as we saw before, that (together with a wage share sufficiently close to one) totally changes the dynamics of this model.

How does the calibration influence the amplification from productivity to unemployment? As seen in (26), the transmission from θ to u depends only on α and on u itself, so there is little disagreement here. The contentious parts of the calibration do not influence this channel. That is, the differences in the amplification of unemployment between the alternative calibrations are

---

[25] Again, we need to remind the reader that our wage elasticity is defined for a one-time permanent change of productivity. Hagedorn and Manovskii (2005) base their analysis on an economy with recurrent and persistent, but not permanent, shocks. Therefore, our calibration results for various parameters can differ somewhat.

inherited from the differences in how these calibrations amplify labor market tightness.

Some Further Remarks on Calibration

What is the value of the labor share? Apparently, relatively minor differences in how close the wage share is to one make a significant difference in the results. It seems to us, however, that wage shares are very difficult to calibrate properly without having the other major input in the model, namely capital. Of course, some search/matching models do allow an explicit role for capital. Pissarides (2000), for example, discusses a matching model where firms, once they have matched, rent capital in a frictionless market for capital. Thus, a neoclassical (or other) production function can be used, and the wage share can be calibrated to the ratio of wage income to total income using the national income accounts. The relevant wage income share, however, is then net of capital income, and the same applies to the definition of output. Alternatively, in Hornstein, Krusell, and Violante (2005), we assume that capital is purchased in competitive markets but that an entrepreneur has to purchase capital first in order to be able to search for a worker—in order to qualify as a "vacant firm." It is an open question as to whether models with capital will also embody a sensitivity of the amplification mechanism to the calibration of the labor share.

What is the value for the wage elasticity? If one insists that wages are less responsive to the cycle than what is implied by Shimer's calibration, then there is more amplification from productivity shocks, and the model's implications are closer to the data. Hall (2005) maintains an even more extreme assumption that wages are entirely rigid; this is why we considered a version of the model with rigid wages. Going back to equation (28), we see that a rather extreme outcome is produced, provided that we still calibrate so that the wage share is close to one. Now inelastic wages and a high wage share interact to boost the amplification mechanism. However, the model has the counterfactual implication that the labor share, w/p, is perfectly negatively correlated with output while only mildly countercyclical in the data.

What is the value for the elasticity of unemployment to benefits? Finally, a possible third clue for calibrating this model can be obtained if one has information about how the economy responds to changes in unemployment compensation.[26] Of course, the absence of controlled experiments makes it difficult to ascertain the magnitude of such effects. The upshot, however, is that if the response of θ to p is large (because b is close to p), then the response of an increase in unemployment compensation would be a very sharp decrease in θ (and increase in unemployment). In particular, as explained in Section 5, the elasticity of the exit rate from unemployment with respect to b equals

---

[26] This way of assessing matching models was proposed in Costain and Reiter (2003).

$(1 - \alpha)$ times $\eta_{\theta b}$. Given $\alpha = 0.72$, the Hagedorn-Manovskii calibration implies that this elasticity equals -6.3 (see Table 3). Thus, a 10 percent rise in unemployment benefits would increase expected unemployment duration $(1/\lambda_w)$ by roughly 60 percent.

The existing estimates of the elasticity of unemployment duration with respect to the generosity of benefits, which are based on "quasi-natural" experiments, are much smaller. Bover, Arellano, and Bentolila (2002) find for Spain that not receiving benefits increases the hazard rate at most by 10 percent, implying a local elasticity of 0.1. For Canada, Fortin, Lacroix, and Drolet (2004) exploit a change in the legislation that led to a rise in benefits by 145 percent for singles below age 30 and estimate an elasticity of the hazard rate around 0.3. For Slovenia, van Ours and Vodopivec (2004) conclude that the 1998 reform which cut benefits by 50 percent was associated with a rise in the unemployment hazard by 30 percent at most, implying an elasticity of 0.6. Finally, an earlier survey by Atkinson and Micklewright (1991) argues that reasonable estimates lie between 0.1 and 1.0.

In sum, these estimates mean that the elasticity implied by the Hagedorn-Manovskii parametrization is between six and sixty times larger than the available estimates.

7. QUANTITATIVE RESULTS FOR THE DIFFERENT CALIBRATIONS

In this section, we show that the three alternative calibrations discussed in Section 6 have very different quantitative implications for the comparative statics discussed in Section 5. Note that, although the values for certain key parameters—β and b in particular—are different, the steady state values of the key aggregate variables are the same across parameterizations. The reason, as explained, is that certain parameters are not uniquely identified in steady state.

Implications for θ, λ_w, and u

Table 3 summarizes the results for the preferred calibrations of Shimer, Hagedorn and Manovskii, and Hall. Recall that Hall's parameterization has a constant wage.

With Shimer's calibration, the model has a very poor amplification mechanism.[27] A 1 percent permanent rise in productivity leads only to a 1.7 percent

---

[27] Though Table 3 contains information about the comparative statics of separation rates, we focus the discussion on the effects of productivity. Shimer (2005) shows that in terms of equation (1), most unemployment volatility in the U.S. economy is accounted for by variations in job creation (the job-finding rate), as opposed to job destruction (the job-separation to rate). Furthermore,

Table 3   Steady State Elasticities

Response of to change in	θ			λ_w			u		
	p	b	σ	p	b	σ	p	b	σ
Shimer	1.72	−0.69	−0.07	0.48	−0.19	−0.02	−0.45	0.18	0.95
Hagedorn & Manovskii	23.72	−22.51	−0.08	6.64	−6.30	−0.02	−6.18	5.87	0.95
Hall	81.70	0.00	−8.17	22.88	0.00	−2.29	−21.30	0.00	3.06

rise in market tightness, a response that is below that in the data by a factor of 16. Similarly, unemployment and the job-finding rate move very little in the wake of a productivity change. Shimer attributes the failure of the model to the fact that, with Nash bargaining, the wage is too closely linked to productivity and absorbs too large a fraction of the productivity fluctuations. As a result, profits do not rise enough to give firms the incentive to create many additional vacancies.

Hall's calibration imposes a constant wage.[28] The consequences of this assumption are striking: Market tightness and unemployment respond almost 50 times more than in Shimer's baseline model. Since wages are fixed, a rise in productivity translates entirely into profits. Firms post many more vacancies, which also boost the volatility of the job-finding rate, λ_w.

Hagedorn and Manovskii's calibration, finally, leads to the best results for the volatility of market tightness and for the job-finding rate with respect to productivity shocks: A 1 percent productivity increase leads to a 20 percent increase of market tightness and a 7 percent increase of the job-finding rate. The main problem, however, is that this calibration induces what seems to be excessive sensitivity of u to unemployment benefits b. The elasticity is about six—almost 20 times larger than the number resulting from Shimer's calibration. To interpret what this magnitude means, consider a policy experiment where unemployment benefits are raised by 15 percent; the unemployment rate would then double under Hagedorn and Manovskii's calibration.[29]

---

as Table 3 demonstrates, variations in the job-separation rate have a negligible effect on the job-finding rate.

[28] For the calibration of Hall's sticky-wage model, we match the wage income share and the unemployment benefits from the Shimer calibration. In all other respects, the calibration is the same as for the Shimer calibration.

[29] The fact that the Hagedorn and Manovskii parameters are chosen such that wages do not respond strongly to changes in productivity implies that wages respond strongly to changes in benefits. For the Hall calibration, wages are simply assumed to be fixed, which imposes no additional restrictions on calibration. Thus, even though wages are less responsive than under Hagedorn and Manovskii, changes in b have no impact on the equilibrium.   When wages are fixed exogenously, the level of benefits is irrelevant.

**Quantitative Implications for the Cyclicality
of the Labor Share**

From equation (32), it is straightforward to rewrite the elasticity of wages, w, and the labor share, s, with respect to a productivity shock, p, as

$$
\begin{aligned}
\hat{w} &= \frac{\beta}{s}\hat{p} + \beta\frac{c\theta/p}{s}\eta_{\theta p}, \text{ and} \\
\hat{s} &= \hat{w} - \hat{p},
\end{aligned}
$$

where $\eta_{\theta p}$ denotes the elasticity of θ with respect to p.

For Hall's calibration, the implications are immediate—the model has the counterfactual implication that the volatility of wages is zero and that the correlation between the labor share and labor productivity is minus one. With Shimer's calibration, $\hat{w} \approx 1.15$, and, hence, wages respond one-for-one to labor productivity, absorbing most of their impact, as explained above. Compared to the data, wages are too volatile. The labor share is essentially acyclical, in contrast with the data. Thus, the baseline calibration of the matching model with a low b also fails along these two dimensions.

Hagedorn and Manovskii's parameter choice is constructed to match $\hat{w} = 0.5$, and therefore $\hat{s} = -0.5$. Under this parameterization, the model is quite successful in matching the elasticity of the labor share, since in the data, the labor share is about as volatile as labor productivity and is countercyclical. Here, it is evident that the choice made by Hagedorn and Manovskii of setting β near zero is useful since one can reconcile a large value for $\eta_{\theta p}$ with small fluctuations in the wage and a countercyclical labor share.

8.   THE MATCHING MODEL WITH AGGREGATE RISK

In the comparative statics exercise above we have studied how long-run outcomes in our model economy respond to one-time permanent changes in parameters. Yet we want to evaluate how well the model matches the business cycle facts of the labor market, and the business cycle is arguably better described by recurrent stochastic changes to parameters. For this reason we now modify the model and include stochastic productivity shocks that are persistent but not permanent.

One might conjecture that the difference between the effects of one-time permanent shocks and persistent—but not permanent—shocks will be smaller, the more persistent the shocks are. In this case the difference between the comparative statics exercise and the analysis of the explicit stochastic model might be small since labor productivity is quite persistent. The autocorrelation coefficient is around 0.8 (see Table 1). It turns out that the difference between the two approaches is noticeable, but it does not overturn the basic conclusion from the comparative statics analysis. If the calibration is such that wages

respond strongly to changes in productivity, then productivity shocks cannot account for the volatility of the labor market.

The modified model can be analyzed in almost closed form—again because free entry makes vacancies adjust immediately to any shock. Thus, as before, unemployment is a state variable, but it will only influence its own dynamics (and, residually, that of vacancies), whereas all other variables will depend only on the exogenous stochastic shocks in the economy. Again, the argument that backs this logic up proceeds by construction: specify an equilibrium of this sort, and show that it satisfies all the equilibrium conditions.

We will focus on a simple case in which the economy switches between a low-productivity state, $p_1 = p(1 - \mu)$, and a high-productivity state, $p_2 = p(1 + \mu)$, with $\mu > 0$. The switching takes place according to a Poisson process with arrival rate τ.[30] The capital values of (un)matched firms and workers, (3) to (6), are easily modified to incorporate the dependence on the aggregate state of the economy:

$$r J_i = p_i - w_i - \sigma (J_i - V_i) + \tau (J_{-i} - J_i), \qquad (34)$$

$$r V_i = -c + \lambda_f (\theta_i) (J_i - V_i) + \tau (V_{-i} - V_i), \qquad (35)$$

$$r W_i = w_i - \sigma (W_i - U_i) + \tau (W_{-i} - W_i), \ \text{ and} \qquad (36)$$

$$r U_i = b + \lambda_w (\theta_i) (W_i - U_i) + \tau (U_{-i} - U_i), \qquad (37)$$

for $i = 1, 2$, where $-i$ denotes 1 if $i = 2$ and vice versa. Each value equation now includes an additional capital gain/loss term associated with a change in the aggregate state. We continue to assume that wages are determined to implement the Nash-bargaining solution for the state-contingent surplus, $S_i = J_i - V_i + W_i - U_i$, and that there is free entry: $V_i = 0$.

We now apply the surplus value definition and the free-entry condition to equations (34) to (37) in the same way as for the steady state analysis in the previous sections. The equilibrium can then be characterized by the following equations:

$$(r + \sigma + \tau) \frac{c}{(1 - \beta) \lambda_f (\theta_i)} = p_i - r U_i + \tau \frac{c}{(1 - \beta) \lambda_f (\theta_{-i})}, \ \text{ and} \quad (38)$$

$$(r + \tau) U_i = b + \frac{\beta \lambda_w (\theta_i) c}{(1 - \beta) \lambda_f (\theta_i)} + \tau U_{-i}, \qquad (39)$$

for $i = 1, 2$. The idea is to see how an increase in μ from zero—when $\mu = 0$, we are formally in the previous model without aggregate shocks—will influence labor market tightness: If p goes up by 1 percent, that is, μ increases by 0.01, by how many percentage points does θ_1 go down and θ_2 go up? And how does the answer depend on τ? We will find answers with two different methods. First we will use a local approximation around $\mu = 0$,

---

[30] The model can easily be extended to include a large but finite number of exogenous aggregate states.

which allows us to derive an elasticity analytically. Then we will look at a particular value of $\tau > 0$ and compute exact values for θ_1 and θ_2.

Local Approximations

For a local approximation at a point where the two states are identical ($\mu = 0$), the equilibrium is symmetric such that θ_1 goes down by the same percentage amount by which θ_2 goes up. For this case, we can show explicitly how the equilibrium elasticity depends on the persistence parameter, τ. We solve for the elasticity in two steps.

First, taking the total derivative of expression (38) with respect to a change in productivity, μ yields

$$\left\{ (r + \sigma + \tau) \frac{\alpha c}{\lambda_{f,i}} \right\} \eta_i = (1 - \beta) \left(\left[(-1)^i p - \frac{\partial r U_i}{\partial \mu} \right] \right) + \left\{ \tau \frac{\alpha c}{\lambda_{f,-i}} \right\} \eta_{-i}, \tag{40}$$

where $\eta_i \equiv (\partial \theta_i / \partial \mu) (1/\theta_i)$ denotes the elasticity of tightness in state i with respect to a change in productivity. Since we consider only a small productivity difference across states, we approximate the terms in curly brackets by the non-state-contingent steady state values for $\mu = 0$. Furthermore, since everything is symmetric, the solution is such that

$$\eta = \eta_2 > 0 > \eta_1 = -\eta. \tag{41}$$

Inspecting the results, it is easy to see that *keeping the effect of productivity on the flow value of unemployment constant*, the absolute value of the elasticity is higher, the lower τ is. The response of labor market tightness to productivity is stronger, the more persistent the shock is. However, higher productivity also raises the flow value of unemployment, which hurts firms, and this effect goes in the opposite direction.

In order to understand the latter effect, we need to solve expression (39) for the flow return on unemployment as a function of labor market tightness:

$$r U_i = b + \frac{\beta}{1 - \beta} c \frac{(r + \tau) \theta_i + \tau \theta_{-i}}{r + 2\tau} \text{ for } i = 1, 2.$$

We see here that if there is no discounting ($r = 0$), productivity would not affect the flow value of unemployment since it would raise θ_2 and lower θ_1, but the two effects are symmetric and cancel each other out. However, discounting results in a larger weight on the current aggregate state. To analyze the effect in detail, take the total derivative with respect to the change in productivity,

**Table 4 Elasticity of Tightness with Respect to Productivity, $\eta_{\theta\,p}$;
Local Approximation**

τ	0.00	0.01	0.02	0.05	0.10	0.50
Average Duration (in years)	∞	25.00	12.50	5.00	2.50	0.50
Shimer	1.72	3.94	5.42	7.04	6.47	2.23
Hagedorn & Manovskii	23.67	29.28	28.24	22.12	15.49	4.42
Hall	81.70	69.32	60.20	43.16	29.33	8.23

μ, to deliver[31]

$$\frac{\partial r U_i}{\partial \mu} = c\theta \frac{\beta}{1 - \beta} \frac{\left[(r + \tau)\,\eta_i + \tau\eta_{-i}\right]}{r + 2\tau} \text{ for } i = 1, 2. \tag{42}$$

Again using symmetry (41) in (42) we obtain

$$\frac{\partial r U_2}{\partial \mu} = -\frac{\partial r U_1}{\partial \mu} = c\theta \frac{\beta}{1 - \beta} \frac{r}{r + 2\tau}\eta. \tag{43}$$

It is apparent that with discounting ($r > 0$), the elasticity of the flow return on unemployment is positively influenced by productivity (it goes up in state two relative to state one), and as shocks become more persistent (as τ falls), the effect is stronger (for a given value of η). We also see that changes in the persistence parameter have a bigger impact on how the flow unemployment value responds to productivity when the persistence parameter is large: τ appears in the denominator so that when it is large, the effects are close to zero. Intuitively, when there is almost no persistence, the flow value of unemployment almost does not react to productivity because it is so short-lived, and small changes in persistence become unimportant too.

Inserting expression (43) into (40) and using symmetry again, one obtains the following expression for the elasticity of labor market tightness:

$$\left\{ (r + \sigma + 2\tau)\frac{\alpha c}{\lambda_f} + \theta\beta\frac{r}{r + 2\tau} \right\} \eta = \frac{(1 - \beta)\,p}{c}. \tag{44}$$

In Table 4, we display the elasticity of labor market tightness with respect to labor productivity for our three different calibrations and how the elasticity depends on the persistence of the aggregate state. For the purpose of business cycle analysis, an average duration of the state between 2.5 ($\tau = 0.1$) and 5 years ($\tau = 0.05$) appears to be appropriate. We see that for business-cycle durations the results differ from the $\tau = 0$ case, which reproduces the numbers from the comparative statics analysis for a one-time permanent shock. In particular, for less-than-permanent shocks, the different calibrations produce

---

[31] We have again approximated the state-contingent values of θ_i with the non-state-contingent steady state value, θ.

Table 5 Elasticity of Tightness with Respect to Productivity; Exact
          Solution for $\mu = 0.005$ and $\tau = 0.05$

	1 to 2	2 to 1
Shimer	7.30	−6.80
Hagedorn & Manovskii	24.98	−19.93
Hall	54.52	−35.28

results that are more similar. The amplification under Shimer's calibration increases relative to the amplification under the alternative calibrations/models. It remains true that with the Shimer calibration labor productivity fluctuations cannot account for the volatility of labor market tightness, whereas the Hagedorn and Manovskii and Hall calibrations come close. Recall that in the U.S. economy labor market, tightness is about 20 times as volatile as labor productivity (see Table 1). For an arrival rate consistent with the persistence of business cycles, $\tau \in [0.05, 0.1]$, labor market tightness for the Shimer calibration is only seven times more volatile than productivity, whereas for the Hagedorn and Manovskii calibrations, tightness is 20 times more volatile than productivity. It is 30 to 40 times as volatile for the Hall calibration.

We also confirm the theoretical analysis above regarding the effects of persistence. The table reveals that the elasticity of labor market tightness with respect to changes in productivity is not necessarily monotone with respect to the arrival rate of the aggregate state change. On the one hand, the more persistent shocks are, the more a productivity increase influences the present value revenue of the firm. As a consequence, more firms enter and labor market tightness goes up. However, as persistence increases, so do the costs of the firms—they are determined by the workers' outside options—and this effect works in the opposite direction. Moreover, as shown above, this latter effect is a particularly important one when persistence is large but a relatively unimportant one when shocks are very short-lived. Thus, as Table 4 shows, the response of labor market tightness to productivity first increases as persistence goes up and then decreases when shocks become close to permanent, and the effect on workers' outside options dominates.

Exact Solution

In Table 5 we display exact results for a case in which a switch from the low productivity to the high productivity represents a 1 percent change in productivity. For Shimer's calibration, this results in a 7.3 percent increase or a 6.8 percent reduction of labor market tightness. The approximation in Table 4 for $\tau = 0.05$ is roughly the average of the elasticities reported in Table 5; thus, the accuracy of the approximations is reasonable.

9.   CONCLUSION: WHERE NEXT?

We have reviewed recent literature that assesses the ability of the search/matching model of the labor market to match some key characteristics of labor markets, namely, the large fluctuations in vacancies and in unemployment. We have, in particular, discussed what features of a calibration seem necessary for matching the data within the context of the standard model or of one augmented with an assumption that real wages are rigid. In this discussion, we have tentatively concluded that there is no wholly satisfactory calibration of the basic setup or a simple alteration thereof that allows the key characteristics of the data to be roughly reproduced. On the one hand, one can assume that the value of being at home is almost as large as that of having a job, but that seems somewhat implausible on a priori grounds, and it implies that there must also be strong sensitivity of unemployment to unemployment benefits, which arguably we do not observe. On the other hand, one can assume rigid wages, but we show that rigid wages necessitate a wage share close to one in order to be powerful in creating large fluctuations in labor market variables, and this route moreover produces an excessively volatile labor share.

It is an open question as to where one might go next. In our view, it seems important to first examine a model with capital, because the results we report above are very sensitive to the value of the labor share. In a model with capital, there is no ambiguity about how one should interpret the labor share. Moreover, a model with capital offers another natural source of fluctuations in vacancies and unemployment, namely, fluctuations in the price of investment goods. Such fluctuations will directly influence the incentives for firms to enter/open new vacancies, and, hence, seem a promising avenue for further inquiry.

REFERENCES

Andolfatto, David. 1996. "Business Cycles and Labor Market Search." *American Economic Review* 86 (1): 112–32.

Atkinson, Anthony B., and John Micklewright. 1991. "Unemployment Compensation and Labor Market Transitions: A Critical Review." *Journal of Economic Literature* 29 (4): 1679–1727.

Bover, Olympia, Manuel Arellano, and Samuel Bentolila. 2002. "Unemployment Duration, Benefit Duration and the Business Cycle." *Economic Journal*, Royal Economic Society 112 (127): 223–65.

Costain, James S. and Michael Reiter. 2003. "Business Cycles,

Unemployment Insurance, and the Calibration of Matching Models."
CESifo Working Paper 1008.

Fisher, Jonas D.M. 2003. "Technology Shocks Matter." Federal Reserve
Bank of Chicago Working Paper 2002-14, rev. 2003.

Fortin, Bernard, Guy Lacroix, and Simon Drolet. 2004. "Welfare Benefits
and the Duration of Welfare Spells: Evidence from a Natural Experiment
in Canada." *Journal of Public Economics* 88 (7–8): 1495–1520.

Greenwood, Jeremy, Zvi Hercowitz, and Per Krusell. 2000. "The Role of
Investment-Specific Technological Change in the Business Cycle."
European Economic Review 44 (1): 91–115.

Hagedorn, Marcus, and Iourii Manovskii. 2005. "The Cyclical Behavior of
Equilibrium Unemployment and Vacancies Revisited." Mimeo,
University of Pennsylvania.

Hall, Robert E. 2005. "Employment Fluctuations with Equilibrium Wage
Stickiness." *American Economic Review* 95 (March): 50–65.

Hornstein, Andreas, Giovanni L. Violante, and Per Krusell. Forthcoming.
"The Replacement Problem in Frictional Economies: A Near
Equivalence Result." *Journal of the European Economic Association.*

Hosios, Arthur J. 1990. "On the Efficiency of Matching and Related Models
of Search and Unemployment." *Review of Economic Studies* 57(2):
279–98.

Lagos, Ricardo. 2000. "An Alternative Approach to Search Frictions."
Journal of Political Economy 108 (October): 851–73.

Merz, Monika. 1995. "Search in the Labor Market and the Real Business
Cycle." *Journal of Monetary Economics* 36 (2): 269–300.

Mortensen, Dale T., and Christopher A. Pissarides. 1994. "Job Creation and
Job Destruction in the Theory of Unemployment." *Review of Economic
Studies* 61 (3): 397–416.

Nash, John F. 1950. "The Bargaining Problem." *Econometrica* 18 (2):
155–62.

Organisation for Economic Co-operation and Development. 1996. *OECD
Employment Outlook.* OECD: Paris.

Osborne, Martin J., and Ariel Rubinstein. 1990. *Bargaining and Markets.*
San Diego: Academic Press.

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

Pissarides, Christopher A. 1985. "Short Run Equilibrium Dynamics of Unemployment, Vacancies, and Real Wages." *American Economic Review* 75 (4): 676–90.

——————. 2000. *Equilibrium Unemployment Theory*, 2nd edition. Cambridge: MIT Press.

Prescott, Edward C. 1986. "Theory Ahead of Business Cycle Measurement." Federal Reserve Bank of Minneapolis *Quarterly Review* 10 (Fall): 9–22.

Rogerson, Richard, Robert Shimer, and Randall Wright. Forthcoming. "Search-Theoretic Models of the Labor Market." *Journal of Economic Literature*.

Shimer, Robert (2004). "The Consequences of Rigid Wages in Search Models." *Journal of the European Economic Association (Papers and Proceedings)* 2 (May–April), 469–79.

——————. 2005. "The Cyclical Behavior of Equilibrium Unemployment and Vacancies." *American Economic Review* 95 (March): 24–49.

van Ours, Jan C., and Milan Vodopivec. 2004. "How Changes in Benefits Entitlement Affect Job-Finding: Lessons from the Slovenian 'Experiment.'" Institute for the Study of Labor Discussion Paper 1181.

Oil Prices and Consumer Spending

Yash P. Mehra and Jon D. Petersen

Although a large body of empirical research indicates that oil price increases have a significant negative effect on real GDP growth, considerable debate exists about both the strength and stability of the relation between oil prices and GDP. In particular, some analysts contend that the estimated linear relations between oil prices and several macroeconomic variables appear much weaker since the 1980s (Hooker 1996).[1]

The evidence of a weakening effect of an oil price change on the macroeconomy in data since the 1980s happens to coincide with another change in the nature of oil price movements: Before 1981, most big oil price movements were positive. Since then, however, oil prices have moved significantly in both directions, reflecting the influences of endogenous macrodevelopments on oil prices. The choppy nature of oil price movements since the 1980s has led some analysts to argue that evidence indicating that oil price changes do not have much of an effect on real GDP is spurious and that the evidence arises from the use of endogenous oil price series. Hamilton (2003), in fact, posits that the relation between oil price changes and real GDP growth is nonlinear, namely, that oil price increases matter but oil price declines do not. Furthermore, oil price increases that occur after a period of stable oil prices matter more than those increases that simply reverse earlier declines. He shows that if the true relation is nonlinear and asymmetric as described above, then the standard linear regression that relates real growth to oil price changes would

[1] Hooker (1996) reports evidence that oil price changes no longer predict many U.S. macroeconomic indicator variables in data after 1973 and that the estimated linear relations between oil price increases and real economic activity indicator variables do appear weaker since the 1980s. Hooker (2002) also reports evidence of weakening of the link between oil prices and inflation since the 1980s.

spuriously appear unstable over a sample period spanning those two sub-periods of different oil price movements.

In order to capture the above-noted hypothesized nonlinear response of GDP growth to oil price changes, Hamilton has proposed a nonlinear transformation of oil price changes. In particular, he uses a filter that weeds out oil price drops and measures increases relative to a reference level, yielding what he calls "net oil price increases."[2] This nonlinear filter, when applied to oil price changes, is supposed to weed out short-term endogenous fluctuations in oil prices, leaving big oil price increases that may reflect the effect of exogenous disruptions to oil supplies. He then shows that the estimated linear relation between net oil price increases and real growth is strong and depicts no evidence of parameter instability over the period 1949 to 2001.[3]

In discussing why oil price shocks have an asymmetric effect on real GDP growth, Hamilton, among others, has emphasized both the "demand-side" and "allocative" channels of influence that oil price shocks have on the real economy. On the demand side, a big disruption in energy supplies has the potential to temporarily disrupt purchases of large-ticket consumption and investment goods that are energy-intensive because it raises uncertainty about both the future price and availability of energy, as in Bernanke (1983).[4] Both households and firms may find it optimal to postpone purchases until they have a better idea of where energy prices are headed after an oil price shock, leading to potential changes in the mix of consumption and investment goods they demand. This postponement and/or shift in the mix of demand may have a nonlinear effect on the economy working through the so-called "allocative" channels that become operative when it is costly to reallocate capital and labor between sectors affected differently by oil price changes. In particular, both

----

[2] Quite simply, his series of net oil price increases is defined as the percentage change from the highest oil price change over the past four, eight, or twelve quarters, if positive, and zero otherwise. This procedure yields net oil price increases measured relative to past one-, two-, and three-year peaks.

[3] Worth noting is that Hamilton (1996, 2003) was not the first to provide evidence of an asymmetric response to oil price increases and oil price declines. Mork (1989) provided evidence indicating that oil price increases had a negative effect on real GNP growth whereas oil price declines did not. However, Hamilton's (2003) paper is the first "rigorous" statistical test of nonlinearity, using flexible functional forms.

[4] The basic argument is that oil price uncertainty may be as important of a determinant of economic activity as the level of oil prices. In case of investment, Bernanke (1983) shows it is optimal for firms to postpone irreversible investment expenditures when they face an increased uncertainty about the future price of oil. When the firm is faced with a choice between adding energy-efficient or energy-inefficient capital, increased uncertainty raises the option value associated with waiting to invest, leading to reduced investment. Hamilton (2003, 366) makes a similar argument for the postponement of purchases of consumer goods which are intensive in the use of energy.

oil price increases and decreases may have a negative effect on GDP growth if oil price effects work primarily through allocative channels.[5]

Of course, oil price increases may affect aggregate spending through other widely known channels. For instance, because oil price increases lead to income transfers from countries that are net importers of oil, such as the United States, to oil-exporting countries, it is plausible for the oil-importing countries to exhibit a reduction in spending. Since an increase in the price of oil would lead to an increase in the overall price level, real money balances held by firms and households would be reduced through familiar monetary channels including the Federal Reserve's counter-inflationary monetary policy response.[6] These income-transfer and real-balance channels, however, imply a symmetric relation between oil price changes and GDP growth.

The asymmetric effect of oil price changes on GDP growth may arise if we consider oil price effects generated through all three channels described above because oil price effects, working through allocative channels, are asymmetric with respect to oil price changes. However, that is not the case for oil price effects working through other channels. Thus, an oil price increase is likely to depress GDP because all three channels (income-transfer, real-balance, and allocative) work to depress aggregate demand in the short run. In contrast, an oil price decline may not stimulate GDP because the positive effect of lower oil prices on aggregate demand generated through the real-balance and income-transfer channels is offset by the negative effect on demand generated through the so-called allocative channels. Another potential contributory factor is the asymmetric response of monetary policy to oil prices—the Federal Reserve tightening policy in response to oil price increases but not pursuing expansionary policies in the face of oil price declines.

This article investigates how much of the negative effect of an oil price increase on real GDP growth works through the consumption channel. As noted above, many analysts have emphasized that big spikes in oil prices affect real growth because they may lead consumers to postpone purchases of large-ticket, energy-guzzling consumption goods. Of course, oil price increases may affect consumer spending, working through other widely known income-transfer and real-balance channels. Another issue investigated here is whether the asymmetric relation between oil prices and real GDP growth found in data holds at the consumption level.

The empirical methodology used to identify the effect of an oil price increase on consumer spending is straightforward: We test for the direct effect of an oil price change on spending that is beyond what can be accounted for by

---

[5] Hamilton (1988) provides a theoretical model in which oil price increases and declines may adversely affect real economic activity because of the high cost of reallocating labor or capital among sectors affected differently by oil price changes.

[6] A good review of these channels appears in Mork (1994).

Figure 1 Quarterly Changes in Oil Prices

Panel A: BEA Prices

Panel B: PPI Prices
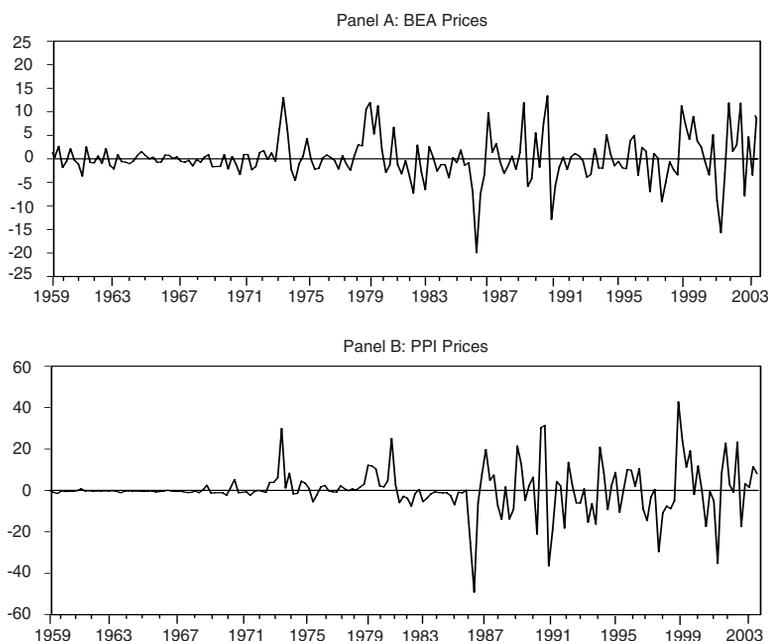
Notes: The BEA oil price series is an index of gas and oil, normalized to 100 in 1982, and deflated. The PPI oil price series is an index of crude oil, normalized to 100 in 1982, deflated, and not seasonally adjusted. The quarterly changes represent the first difference of the log of the prices, multiplied by 100.

other economic determinants of spending, such as households' labor income and net worth. We alternatively measure oil price shocks as "positive oil price increases" (Mork 1989) or "net oil price increases" (Hamilton 1996, 2003). The sample period studied is 1959:Q1 to 2004:Q2.

The empirical work presented here finds evidence of a nonlinear relation between oil price changes and growth in real consumer spending: Oil price increases have a negative effect on spending whereas oil price declines have no effect. The estimated negative effect of an oil price increase on consumer spending is large if oil price increases are measured as net increases, suggesting oil price increases that occur after a period of stable oil prices matter more than those increases that simply reverse earlier declines. Furthermore, the estimated negative effect on spending is also large if consumer spending is broadly defined to include spending on durable goods, suggesting the possible negative influence of oil price increases on the purchase of big-ticket consumption goods. Finally, the estimated oil price coefficients in

the consumption equation do not show parameter instability during the 1980s, the period when oil prices moved widely for the first time in both directions.

This article is organized as follows. Section 1 examines the behavior of two oil price series to highlight the choppy nature of oil price changes since 1981 and to derive estimates of oil shocks as defined in Hamilton (1996, 2003). Section 2 presents the aggregate empirical consumer spending equation that underlies the empirical work here and reviews theory about how oil price shocks may affect the macroeconomy. Section 3 presents the empirical results, and Section 4 contains concluding observations.

1. A PRELIMINARY REVIEW OF OIL PRICE CHANGES AND NET OIL PRICE INCREASES

In this section we first examine the behavior of two oil price series and then review the rationale behind the construction of net oil price increases as measures of oil price shocks, as in Hamilton (1996, 2003). The first series, prepared by the Bureau of Economic Analysis (BEA), measures gas and oil prices paid by consumers. The second series is the Producer Price Index (PPI) for crude petroleum prepared by the Bureau of Labor Statistics (BLS). In estimating the impact of oil price increases on real GDP growth, analysts have commonly focused on the oil price series for crude petroleum. We, however, focus on the consumer oil price series because changes in consumer spending are likely to be correlated with changes in oil prices actually faced by consumers rather than with changes in the producer price of crude petroleum.

Figure 1 plots the first differences of logs of these two oil price indexes. (The reported differences are multiplied by 100.) This figure highlights one key change in the time-series behavior of oil price changes over 1959 to 2004: Before 1981, big oil price movements were mostly positive. Since then, however, oil prices have moved widely in both directions. Hamilton argues that this change in the time-series behavior of oil price changes reflects the growing influence of endogenous macroeconomic developments on oil prices, namely that oil prices during the 1980s had been influenced dramatically by demand conditions. As a result of the increased endogenous nature of oil price movements, the estimated linear relation between oil price changes and real GDP growth appears unstable over the sample period that includes pre- and post-1980s oil price changes.

Hamilton proposes a nonlinear transformation of oil price changes in order to uncover the relation between the exogenous oil price movements and GDP growth. As indicated at the outset, he uses a filter that leaves out oil price declines and measures increases relative to a reference level, yielding what he calls net oil price increases. Briefly, a net oil price increase series is the percentage change from the highest oil price reached over the past four, eight, or twelve quarters, if positive, and zero otherwise.

Figure 2 Using BEA's Oil Price Index

Panel A: Changes in Oil Prices

Panel B: Net Oil Price Changes (1-year Horizon)

Panel C: Net Oil Price Changes (2-year Horizon)
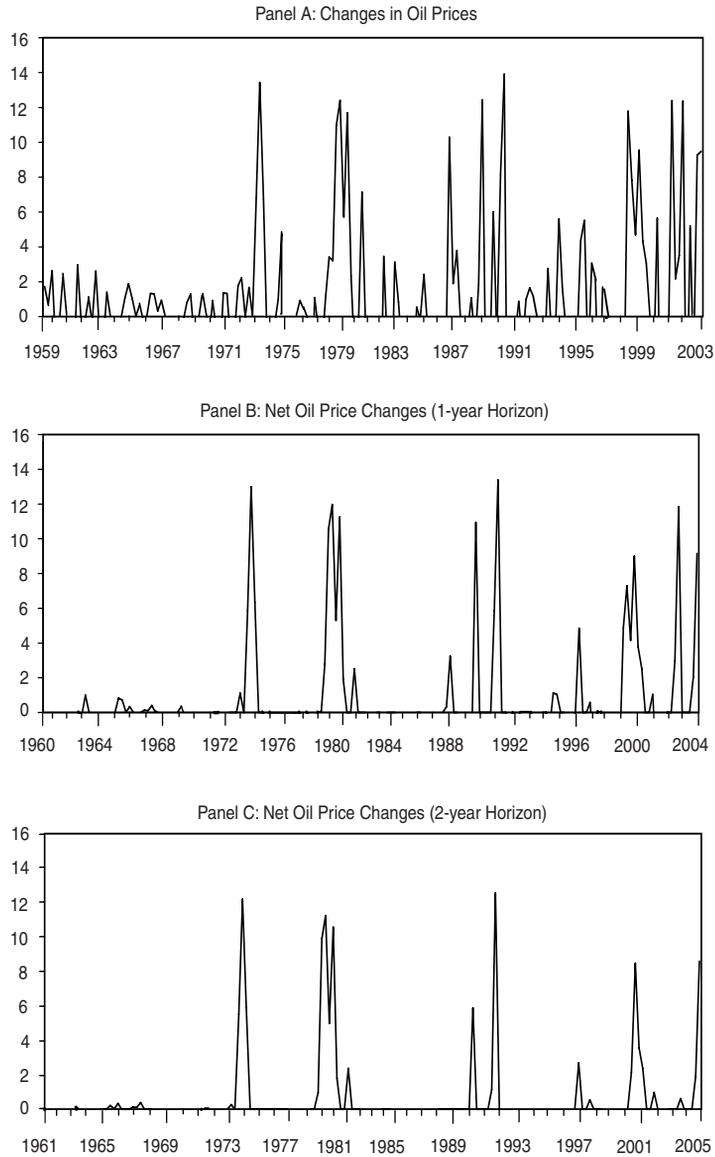
Notes: The oil price index is for the BEA's index of gas and oil, and it is deflated. The net oil price increase is the maximum of (a) zero and (b) the difference between the log level of the oil price index for quarter t and the maximum value for the level achieved during the previous four (Panel B) or eight (Panel C) quarters.
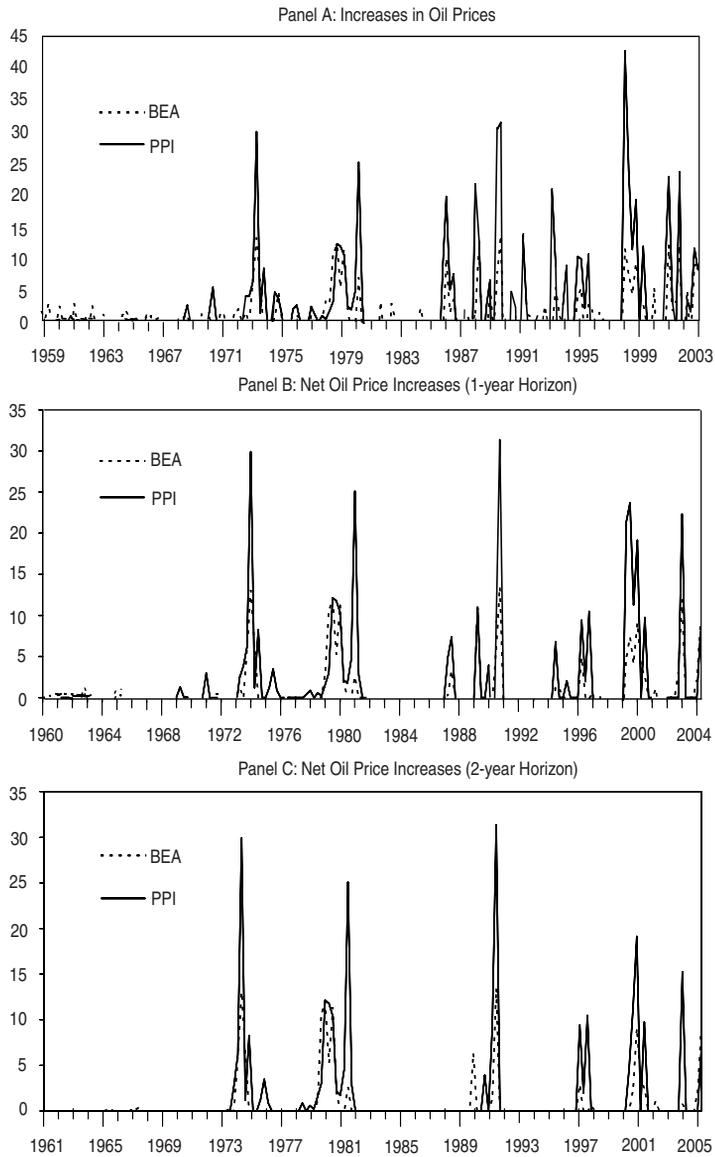
Figure 2 plots oil price increases using the consumer oil price series. Panel A of Figure 2 plots only quarterly increases, whereas Panels B and C plot net oil price increases measured relative to past one- and two-year peaks, respectively. If we compare Panels A, B, and C, we may note that the use of a nonlinear filter results in weeding out certain increases in oil prices that were simply corrections to earlier declines. For example, the big spike in oil prices observed during the first quarter of 2003 does not show up in the net oil price increases measured relative to two-year peaks because it followed the big decline of oil prices in 2001. If we focus on net oil price increases measured over two-year peaks, we get relatively few episodes of oil price spikes, occurring in 1973–1974, 1979–1980, 1990, 1999–2000, and 2004. Hamilton argues that these oil price spikes can be attributed to disruptions in oil supplies associated with military conflicts and, hence, exogenous to the U.S. economy, with one exception.[7] The most recent spike in oil prices may be attributed mainly to the surge in world oil demand (Hamilton 2004).

Figure 3 plots net oil price increases using both oil price series. Two observations stand out. The first is that the net oil price increase series for crude petroleum gives qualitatively similar inferences about the nature of oil price movements as does the consumer price series for gas and oil. However, net oil price increases measured using the consumer oil price series are significantly smaller than those derived using the producer price of crude petroleum. The empirical work presented below uses the net oil price increases created using the consumer oil price series.

2. EMPIRICAL AGGREGATE CONSUMER SPENDING EQUATIONS

The empirical strategy used to identify the consumption effect on an oil price increase is to look for the direct impact of a "net oil price increase" on consumer spending beyond that which can be accounted for by other economic determinants of consumption. We use as control variables economic determinants suggested by the empirical "life-cycle" aggregate consumption equations estimated in Mehra (2001). The empirical work in Mehra (2001) identifies income and wealth as the major economic determinants of consumer spending, and the "life-cycle" aggregate consumption equations provide sensible estimates of income and wealth elasticities, besides predicting reasonably well the short-term behavior of consumer spending. In particular, the empirical short-term consumption equation used here is based on the following consumption equations:

---

[7] The dates of military conflicts that led to declines in world production of oil are November 1973 (Arab-Israel War), November 1978 (Iranian Revolution), October 1980 (Iran-Iraq War), and August 1990 (Persian Gulf War). See Hamilton (2003, 390).

Figure 3 Comparison of Oil Price Increases

Panel A: Increases in Oil Prices

Panel B: Net Oil Price Increases (1-year Horizon)

Panel C: Net Oil Price Increases (2-year Horizon)

Notes: The oil price series are identical to those of Figures 1 and 2.

$$C_t^p = a_0 + a_1 Y_t + a_2 Y_{t+k}^e + a_3 W_t, \text{ and} \qquad (1)$$

$$\Delta C_t = b_0 + b_1 \left(C_{t-1}^p - C_{t-1} \right) + b_2 \Delta C_t^p + \sum_{s=1}^{k} b_{3s} \Delta C_{t-s} + \mu_t, \qquad (2)$$

where C_t^p is planned current consumption, C_t is actual current consumption, Y_t is actual current-period labor income, W_t is actual current-period wealth, and Y_{t+k}^e is average anticipated future labor income over the earning span (k) of the working-age population.

Equation 1 simply states that aggregate planned consumption depends upon the anticipated value of lifetime resources, which equals current and anticipated future labor income and current value of financial assets. This equation identifies income and wealth as the main economic determinants of aggregate planned consumption.

Equation 2 allows for the possibility that actual consumption in a given period may not equal planned consumption, reflecting the presence of adjustment lags and/or habit persistence. In this specification, changes in current-period consumption depend upon changes in current-period planned consumption, the gap between last period's planned and actual consumption, and lagged actual consumption. The disturbance term μ in (2) captures the short-run influences of unanticipated shocks to actual consumer spending. If we substitute (1) into (2), we get the short-run dynamic consumption equation (3):

$$
\begin{aligned}
\Delta C_t \;=\;& b_0 + b_1 (C_{t-1}^p - C_{t-1}) + b_2 (a_1 \Delta Y_t + a_2 \Delta Y_{t+k}^e + a_3 \Delta W_t) \\
& + \sum_{s=1}^{k} b_{3s} \Delta C_{t-s} + \mu_t.
\end{aligned}
\qquad (3)
$$

The key feature of equation (3) is that changes in current-period consumption depend upon changes in income and wealth variables, besides depending upon the last period's gap between the level of actual and planned consumption.

We estimate the "direct" influence of oil price changes on consumer spending by including lagged values of net oil price increases in the short-term consumption equation (3). As another control variable, we also include lagged values of changes in the nominal federal funds rate in order to capture the possible additional influence of changes in short-term interest rates on consumer spending. The inclusion of a short-term nominal interest rate in the consumption equation also controls for the potential influence of oil price

increases on spending that work through the monetary policy channel, arising as a result of the Federal Reserve's monetary policy response to oil shocks.[8]

The empirical work below makes two additional assumptions. The first is that expected future labor income is simply proportional to expected current labor income. The second assumption is that current-period values of income and wealth variables are not observed, and, hence, planned consumption depends upon their known past values. Under these assumptions, the estimated short-consumption equation is

$$\Delta C_t = \beta_0 + \beta_1 (C_{t-1}^p - C_{t-1}) + \beta_2 \Delta Y_{t-1} + \beta_3 \Delta W_{t-1} \qquad (4)$$

$$+ \sum_{s=1}^{6} \beta_{4s} \Delta C_{t-s} + \sum_{s=1}^{3} \beta_{5s} \Delta Oil\,\mathrm{Pr}\,ices_{t-s} + \sum_{s=1}^{3} \beta_{6s} \Delta FR_{t-s},$$

where

$$C_t^p = \alpha_0 + \alpha_1 Y_t + \alpha_2 W_t.$$

In the empirical, short-term consumption equation (4), changes in current consumer spending depend on lagged values of changes in income, net worth, the short-term nominal interest rate, and oil prices, besides depending on lagged changes in consumption and the gap between the level of actual and planned consumption.

3. OIL PRICE EFFECT CHANNELS AND THE REDUCED-FORM EMPIRICAL EVIDENCE

In this section, we review theory on how oil price increases may affect the real economy and discuss its implications for interpreting the evidence of a relation between oil price changes and consumer spending found using the aggregate consumer spending equation (4).

How do oil prices, in theory, affect the macroeconomy? A simple answer is that previous research does not offer any dominant theoretical mechanism.[9] Researchers have emphasized several different theoretical mechanisms through which oil may affect the macroeconomy. One of those mechanisms focuses on the inflation effect of oil price increases and its associated consequences that work through the so-called real-balance and monetary policy

---

[8] A debate exists about whether the contractionary consequences of oil price shocks are due to oil price shocks themselves or to the monetary policy that responds to them. The evidence so far is not very conclusive. See, for example, Leduc and Sill (2004) who investigate this question in a calibrated general equilibrium model in which oil use is tied to capital utilization. Their findings suggest that while the monetary policy rule in place can contribute to the magnitude of the negative output response to an oil-price shock, the "direct" effect of the oil-price increase is the more important factor.

[9] See Hooker (2002), Hamilton (2003), and references cited in both.

channels. The real-balance channel posits that oil price increases lead to inflation, lowering real money balances held by the households and firms in the economy and thereby depressing aggregate demand through familiar monetary channels. The monetary policy channel becomes operative if the Federal Reserve tightens policy in response to inflation induced by oil prices, which may exacerbate further the negative output effect associated with oil shocks.

Another theory of how oil may affect the macroeconomy arises out of viewing an oil price as an import price. In particular, oil price increases lead to income transfers from countries that are net importers of oil, such as the United States, to oil-exporting countries. The first-round effect of this reduction in income is to cause economic agents in oil-importing countries to reduce their spending, leading to reduced aggregate demand. [10]

Some other channels through which oil may affect the macroeconomy arise when oil is modeled as another input in the production function. If oil and capital are complements in the production process, then oil price increases lead to a decline in the economy's productive capacity as agents respond to higher oil prices by reducing their utilization of both oil and capital. In this case, oil price increases lead to negative transitional output growth as the economy moves to a new steady-state equilibrium growth path. To the extent oil price increases raise uncertainty about both its future price and availability, oil price increases may also lead to the postponement of purchases of large-ticket consumption and investment goods, as in Bernanke (1983).[11] Hence, oil price increases have the potential to affect real growth by reducing both potential output and aggregate demand.

Another theoretical mechanism that links oil to the macroeconomy has emphasized the allocative effects of oil price shocks (Hamilton 1988, 2003). An oil price increase is likely to reduce demand for some goods but possibly raise demand for some others. For example, demand for inputs is likely to fall in sectors that use energy but likely to increase in sectors that produce energy. If it is costly to reallocate capital or labor between sectors affected differently by an oil price increase, then aggregate employment and output will decline in the short run. In this framework, an oil price decrease may also lower demand for some goods (demand for inputs used in energy-producing sector) and, hence, may be contractionary if labor or capital could not be moved to favorably affected sectors.

---

[10] The second-round effects arise from, among others, the recycling of income transfers, which is increased income of oil-exporting countries that leads to increased demand for products of the oil-importing countries, thereby offsetting the initial fall in aggregate demand. A recent empirical study, however, finds that among most oil importing countries, including the United States, oil price increases have a negative effect on economic activity (Jimenez-Rodriguez and Sanchez 2004).

[11] See footnote 4.

The discussion above implies that oil price increases may, in theory, affect real growth through several different channels, as emphasized by different researchers. This review then raises another question: Does the empirical evidence reported in previous research support any dominant theoretical mechanism? The answer to this question again appears to be "no" because most of the empirical evidence is based on estimated reduced-form regressions that relate changes in GDP growth to changes in oil prices, controlling for the influences of some other variables on real growth such as lagged real GDP growth, short-term interest rate, import price inflation, etc. As is well known, the evidence based on reduced-form regressions indicating that oil price increases have a significant effect on the macroeconomy may be consistent with several different theoretical mechanisms.

However, analysts who have reported the empirical evidence of the nonlinear and asymmetric relation between oil prices changes and real GDP growth assert that such evidence does appear to favor mechanisms in which oil shocks affect real GDP through the so-called uncertainty and allocative channels, as in Hamilton (2003). The main reason for the emphasis on allocative channels is that other channels, such as income-transfer and real-balance, imply a symmetric relationship between oil price changes and GDP growth. The asymmetry may arise because oil price effects that work through allocative channels differ from those that work through other channels already mentioned. Thus, an oil price increase is likely to depress GDP because all three channels described above (income-transfer, real-balance, and allocative) work to depress aggregate demand. In contrast, an oil price decline may not stimulate GDP because the positive effect of lower oil prices on aggregate demand generated through the real-balance and income-transfer channels is offset by the negative effect on demand generated through the so-called allocative channels. Another factor that may augment the asymmetric response of oil prices to GDP is the asymmetric response of monetary policy to oil prices—the Federal Reserve tightening policy in response to oil price increases but not pursuing expansionary policies in face of oil price declines.[12]

Given the considerations noted above, we investigate whether oil price increases directly affect consumer spending and whether the nonlinear and asymmetric relation between oil prices and real GDP found in previous research hold at the consumption level.

---

[12] Some analysts have argued that during the 1980s and 1990s the Federal Reserve followed an "opportunistic" disinflation policy in the sense that if actual inflation declined due to some shocks, the Federal Reserve lowered its inflation target and adjusted policy to maintain the lower inflation rate. Since oil price shocks have been an important source of movements in inflation, the Federal Reserve following an opportunistic disinflation policy may not follow an expansionary policy if actual inflation falls below its short-term target in response to an oil price decrease. In that regime, a relatively tight policy offsets the expansionary effect of an oil price decrease on the real economy. The quantitative importance of this oil-price policy interaction channel remains a subject of future research.

4.   EMPIRICAL RESULTS

In this section, we present and discuss the evidence regarding the effect of oil price changes on consumer spending, using estimated reduced-form consumer spending equations as shown in (4). The consumption equations are estimated using quarterly data over 1962:Q1 to 2004:Q2 and measurement of variables as in Mehra (2001).[13]

Estimates of Oil Price Effects

Table 1 reports coefficients from the short-term consumption equation (4) estimated using total consumer spending and three different measures of oil price changes: quarterly oil price changes, positive increases in oil price, and net oil price increases. We report the sum of coefficients that appear on the oil price variable and the t-value for a test of the null hypothesis that the sum of oil price coefficients is zero. Since the consumption equation is estimated including lagged consumption, the cumulative response of spending to an oil price increase is likely to differ from its short-term response. Hence, we also report the cumulative size of the coefficient that appears on the oil price variable, which is just the short-term oil price coefficient divided by one minus the sum of estimated coefficients on lagged consumption. We also report estimated coefficients on other control variables that appear in the short-term consumption equation, including lagged consumption, labor income, household net worth, and the short-term interest rate.

The columns labeled (1) through (5) in Table 1 contain coefficients from the short-term consumption equation estimated using different measures of oil price changes. The estimates with quarterly oil price changes are in column (1), those with positive oil price changes are in column (2), and those with net oil price increases measured relative to one-, two-, and three-year peaks are in columns (3), (4), and (5), respectively. If we focus on the oil price coefficient, the estimated coefficient on the oil price variable has a negative sign and is statistically different from zero only when oil price changes are measured either as oil price increases or net oil price increases (compare t-values on the oil price change variable in different columns of Table 1). The estimated coefficient on the quarterly oil price change variable is small and not statistically different from zero. The small t-value of the null hypothesis that the estimated coefficient on oil price declines when added into the short-term consumption equation containing oil price increases, given in column (2), suggests that oil

---

[13] Consumption is measured as per capita consumption of durables, nondurables, and services in 2000 dollars (C). Labor income is measured as per capita disposable labor income, in 2000 dollars (Y). Household wealth is measured as per capita household net worth in 2000 dollars. The short-term interest rate is the nominal federal funds rate. The oil price series measures gas and oil prices paid by consumers, prepared by the BEA.

Table 1 Empirical Aggregate Consumer Spending Equations

$$c_t = \beta_0 + \beta_1 \left(c_{t-1} - c_{t-1}^p\right) + \beta_2 \Delta y_{t-1} + \beta_3 \Delta w_{t-1} +$$

$$\sum_{s=1}^{6} \beta_{4s} \Delta c_{t-s} + \sum_{s=1}^{3} \beta_{5s} \Delta oilprices_{t-s} + \sum_{s=1}^{3} \beta_{6s} \Delta FR_{t-s}$$

$$\text{where } c_t^p = f_0 + f_1 y_t + f_2 w_t + f_3 TR_t$$

Independent Variables	(1)	(2)	(3)	(4)	(5)
Δc_{t-s}	0.660 (4.6)	0.560 (4.2)	0.580 (4.5)	0.540 (4.2)	0.530 (4.0)
Δy_{t-1}	0.110 (2.4)	0.120 (2.5)	0.100 (2.1)	0.100 (2.0)	0.100 (2.1)
$\Delta w1$	0.050 (2.5)	0.050 (2.9)	0.040 (2.6)	0.040 (2.6)	0.040 (2.3)
ΔFR_{t-s}	−0.003 (4.5)	−0.003 (4.3)	−0.003 (4.3)	−0.003 (4.1)	−0.003 (4.1)
$c_{t-1} - c_{t-1}^p$	−0.130 (3.3)	−0.120 (3.0)	−0.120 (3.1)	−0.130 (3.3)	−0.130 (3.2)
Δoil_{t-s}	−0.100 (0.4)				
$P\Delta oil_{t-s}$		−0.030 (1.6)			
$NP\Delta oil_{t-s}$					
1-year			−0.050 (1.8)		
2-year				−0.070 (2.1)	
3-year					−0.070 (2.1)
Adj.R^2	0.3600	0.3800	0.3700	0.3800	0.3800
SER	0.0055	0.0054	0.0054	0.0053	0.0053
Cumulative oil price coefficient	−0.0200	−0.0800	−0.1200	−0.1600	−0.1600

Notes: The coefficients (t-values in parentheses) reported above are ordinary least squares estimates of the short-term consumption equation. Δc is change in real consumer spending, Δy is change in labor income, Δw is change in net worth, ΔFR is change in the nominal federal funds rate, c^p is planned consumption, Δoil is change in oil prices, $P\Delta oil$ is positive changes in oil prices, $NP\Delta oil$ is net oil price increases measured relative to one-, two-, and three-year peaks, Adj.R^2 is the adjusted R-squared, and SER is the standard error of regression.

The coefficient reported on Δc_{t-s} is the sum of coefficients that appear on six lagged values of consumer spending and the coefficient on the oil price variable is the sum of coefficients that appear on three lagged values of the oil price variable. The cumulative oil price coefficient is the coefficient on lagged oil divided by one minus the coefficient on lagged consumption. The effective sample period is 1961:Q1 to 2004:Q2.

price declines have no effect on consumer spending. Together these estimates suggest only oil price increases have a negative effect on consumer spending, implying the presence of an asymmetric relation between oil price changes and consumer spending.

The estimated size of the cumulative oil price response coefficient is −0.08 when oil price changes are measured as oil price increases and ranges between −0.12 to −0.16 when oil price changes are measured as net oil price increases.

Those estimates imply that a 10 percent increase in oil prices is associated with the level of consumer spending at the end of six quarters being anywhere between 0.80 percent to 1.60 percent lower than what it otherwise would be. This effect includes the direct effect of the net oil price increase and the indirect effect that comes through lagged consumption. Given that consumer spending is two-thirds of GDP, the above estimates imply that a 10 percent increase in the price of oil working through the consumption channel will be associated with the level of GDP that is anywhere between one-half to one percentage point lower than what it otherwise would be. In Hamilton (2003), a 10 percent increase in the price of oil is associated with the level of GDP that is 1.4 percent lower than what it otherwise would be, which is above the estimated range, suggesting oil price increases may also affect real GDP working through investment and other components of aggregate demand.

It is worth pointing out that estimated coefficients on other variables such as household labor income, net wealth, and changes in the short-term nominal interest rate have theoretically correct signs and are statistically different from zero (see t-values for those variables in different columns in Table 1). Furthermore, the estimated coefficient on the so-called error-correction variable, which measures the effect on current spending of last period's gap between actual and planned spending, as in (4), is correctly negatively signed and statistically different from zero.

Table 2 presents some robustness analysis of oil price effects with respect to few changes in the specification of the aggregate consumer spending equation. The estimates of oil price effects discussed above are derived using consumer spending that includes spending on durable goods because oil price shocks are hypothesized to affect spending on big-ticket consumer goods that are intensive in the use of energy. But since oil price increases may affect consumer spending by working through other channels, we also estimate the short-term consumption equations that include spending only on nondurable goods and services. Furthermore, we also estimate the aggregate consumer spending equation without controlling for the direct effect of changes in the short-term nominal interest rate on spending. Many analysts have argued that the negative effect of oil price shocks observed on real GDP growth may be due not to oil price shocks themselves but to the monetary policy response to them. Although this issue can not be examined in a rigorous manner using reduced-form spending equations, we offer some preliminary evidence by examining whether the magnitude of oil price effects on consumer spending is sensitive to the exclusion of the interest rate variable.

Table 2 reports estimates of the cumulative oil price coefficient found using consumer spending on nondurable goods and services with and without the interest rate. It also includes results of total consumer spending. Three observations stand out: The first is that the estimated negative effect of an oil price increase on consumer spending is large if spending is broadly defined to

Table 2 Sensitivity Analysis

	Cumulative Oil Price Coefficient			
Measures of Consumer Spending	$P\Delta oil$	1-year	$NP\Delta oil$ 2-year	3-year
Consumer spending including durables				
with ΔFR_{t-s}	−0.08*	−0.12*	−0.16*	−0.16*
without ΔFR_{t-s}	−0.09*	−0.13*	−0.18*	−0.17*
Consumer spending without durables				
with ΔFR_{t-s}	−0.03	−0.05	−0.09*	−0.08*
without ΔFR_{t-s}	−0.04	−0.06	−0.09*	−0.09*

Notes: See notes in Table 1.

* significant at the 0.05 level

include spending on durable goods (compare the size of the oil price coefficient estimated using alternative measures of spending with and without spending on durables, as shown in Table 2). The second observation is that the magnitude of the oil price effect on spending estimated here is not overly sensitive to the exclusion of the interest rate variable from the short-term consumption equation. The third point to note is that the estimated negative effect on spending of net oil price increases is larger than that of positive increases in oil prices, suggesting those increases that occur after a period of stable oil prices affect spending more than oil price increases that simply reverse earlier declines (compare the relative magnitude of the oil price coefficient on oil price increases and net oil price increases, as shown in Table 2). Together these results are consistent with the view that oil price increases affect spending by influencing spending on durable goods and that oil price increases have a direct effect on spending that is beyond what could occur through the monetary policy response to oil prices.

Stability of Oil Price Coefficients

Hamilton (2003) has argued that if we focus on exogenous oil price increases, then the estimated linear relation between exogenous oil price shocks and real GDP growth remains stable. We follow Hamilton in measuring exogenous oil price shocks as net oil price increases believed to be associated with major disruptions to world oil supplies. We now examine whether such a result holds at the consumption level. As indicated before, oil prices have swung widely in both directions since 1981. Hence, we investigate whether oil price coefficients in the aggregate consumer spending equation (4) have changed since 1981.

We implement the test of stability of oil price coefficients using a dummy variable approach with the break date around 1981. We also implement the stability test treating the break date unknown in the 1980s. In particular, consider the following aggregate consumption equation:

$$\Delta C_t = \beta_0 + \sum_{s=1}^{3} \beta_{1s} \Delta OilPrices_{t-s} + \sum_{s=1}^{3} d_{1s}(\Delta OilPrices^* DU)_{t-s}$$
$$+\beta_{2s} X_{t-s} + \varepsilon_t, \tag{5}$$

where DU is a dummy variable, defined as unity over the period since the break date and zero otherwise; X is the set of other control variables including lagged values of consumer spending, labor income, household net worth, and changes in the nominal interest rate, as in (4). In (5), the test of the null hypothesis of stable oil price coefficients against the alternative that they have changed at date t_1 is that all slope dummy coefficients are zero, i.e., $d_{1s} = 0$, $s = 1, 2, 3$. Under this null hypothesis, the standard F statistic F_{t1} would have a chi-squared distribution with three degrees of freedom, $\chi^2(3)$, asymptotically.[14]

We calculate the value of the statistic for every possible value of the break date between 1981:Q1 to 1990:Q4, using oil price increases and net oil price increases as alternative measures of oil price changes. Panel A in Figure 4 plots the p-value from this test as a function of the break date t_1 using oil price increases, whereas panels B through D do so using net oil price increases. As can be seen, the p-value from this test is above the 0.05 p-value for all the break dates and for all measures of oil price increases. These test results suggest that the nonlinear relations between oil price changes and growth in consumer spending do not depict any parameter instability during the 1980s.[15]

5.   CONCLUDING OBSERVATIONS

This article reports empirical evidence indicating that oil price increases have a negative effect on consumer spending whereas oil price declines do not. Furthermore, oil price increases that occur after a period of stable oil prices matter more than oil price increases that reverse earlier declines. This finding of a

---

[14] The aggregate consumption equations have been estimated allowing for the presence of a heteroscedastic disturbance term, and, hence, the standard F statistic has a chi-squared, not F, distribution.

[15] The inference regarding stability of oil price coefficients does not change if we were to treat the break date from 1981:Q1 to 1990:Q4 as unknown and compare the largest value of the F statistic over possible break dates with the 5 percent critical value, as in Andrews (1993). The largest value of the F statistic is 4.7 when oil price changes are measured as oil price increases, which is below the 5 percent critical value of 9.29 given in Andrews (1993, Table 1, using $\pi = 0.48$, $p = 3$ restrictions). The largest F takes values 6.1, 5.2, and 4.9 for net oil price increases measured relative to one-, two-, and three-year peaks, respectively. For these alternative measures of oil price changes, the largest F remains below the 5 percent critical value, suggesting that oil price coefficients do not depict any parameter instability during the 1980s.

Figure 4 Chow Test for Stability of Oil Price Coefficients

Panel A: Oil Price Increases

Panel B: Net Oil Price Increases (1-year)

Panel C: Net Oil Price Increases (2-Year)

Panel D: Net Oil Price Increases (3-Year)
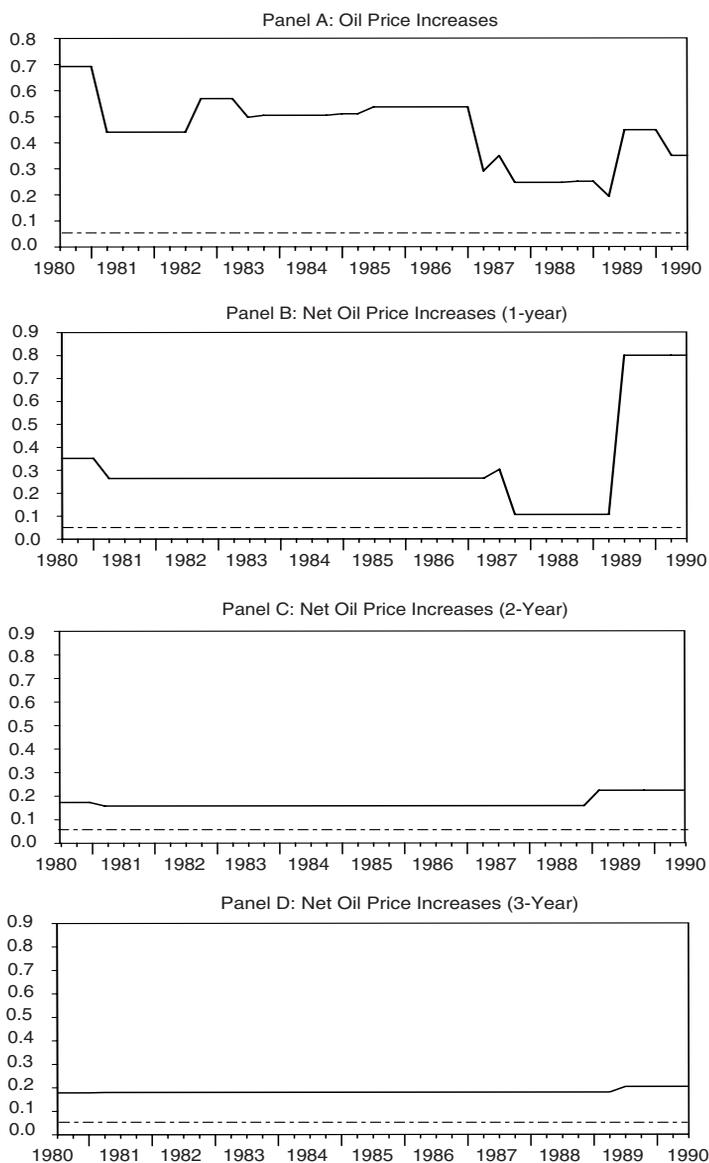
Notes: Each figure plots the p-value of a Chow test where the null hypothesis is that oil price coefficients are stable against an alternative that they have changed at the given date. The dashed lines indicate a p-value of 0.05.

nonlinear and asymmetric relation between oil price changes and consumer spending is in line with what other analysts have found existing between oil price changes and aggregate real economic activity such as real GDP growth.

The results reported here also indicate that oil price increases have a stronger effect on consumer spending if spending is broadly defined to include spending on durables, suggesting oil price increases may be affecting consumer spending by affecting demand for consumer durable goods. However, oil price increases may be affecting consumer spending by working through other channels as well because oil price increases continue to have a significant effect if spending includes only nondurables and services.

The evidence indicating that oil price decreases have no effect on consumer spending is derived using reduced-form consumer spending equations and, hence, may be consistent with several different theoretical mechanisms. One explanation of why an oil price decrease does not have a significant effect on spending is that the positive effect of an oil price decrease generated through the real-balance and income-transfer channels offsets the negative effect on spending generated through allocative channels. Furthermore, if the Federal Reserve does not lower the funds rate in response to oil price declines but raises it in response to oil price increases, we may also find that oil price decreases have no significant effect on spending whereas oil price increases do. Without help from a structural model, we cannot determine which of these two mechanisms is dominant in generating the asymmetry found in data.

The empirical work here focuses on the effect of "exogenous" oil price increases (measured by net oil price increases) on consumer spending, namely, oil price increases caused by exogenous events such as those resulting from disruptions to oil supplies caused by military conflicts. However, increases in oil prices that are due to a rising world demand for oil may not necessarily raise uncertainty about future energy supplies and prices and thus may not adversely affect demand for durable consumption goods, as emphasized in this literature. To the extent that oil price increases affect spending by working through other channels, however, oil price increases, even if due to rising world oil demand, could still adversely affect consumer spending in the short run.

REFERENCES

Andrews, Donald W. K. 1993. "Tests for Parameter Instability and Structural Change with Unknown Change Point." *Econometrica* 16 (4): 821–56.

Bernanke, B.S. 1983. "Irreversibility, Uncertainty and Cyclical Investment." *Quarterly Journal of Economics* 97 (1): 86–106.

Hamilton, James D. 2004. "Causes and Consequences of the Oil Shock of 2004." Available at: http://weber.ucsd.edu/~jhamilto/Oil_Aug04.htm (accessed 20 May 2005).

——————————. 2003. "What is an Oil Shock?" *Journal of Econometrics* 113 (2): 363–98.

——————————. 1996. "This is What Happened to the Oil Price-Macroeconomy Relationship." *Journal of Monetary Economics* 38 (2): 215–20.

——————————. 1988. "A Neoclassical Model of Unemployment and the Business Cycle." *Journal of Political Economy* 96 (3): 593–617.

Hooker, M. A. 1996a. "What Happened to the Oil Price-Macroeconomy Relationship?" *Journal of Monetary Economics* 38 (2): 195–213.

——————————. 1996b. "This is What Happened to the Oil Price-Macroeconomy Relationship: Reply." *Journal of Monetary Economics* 38 (2): 221–22.

——————————. 2002. "Are Oil Shocks Inflationary? Asymmetric and Nonlinear Specifications versus Changes in Regime." *Journal of Money, Credit and Banking* 34 (2): 541–61.

Leduc, Sylvain, and Keith Sill. 2004. "A Quantitative Analysis of Oil-Price Shocks, Systematic Monetary Policy, and Economic Downturns." *Journal of Monetary Economics* 51 (4): 781–808.

Limenez-Rodriguez, Rebeca, and Marcelo Sanchez. 2004. "Oil Price Shocks and Real GDP Growth." European Central Bank Working Paper No. 362.

Mehra, Yash P. 2001. "The Wealth Effect in Empirical Life-Cycle Aggregate Consumption Equations." Federal Reserve Bank of Richmond *Economic Quarterly* 87 (2): 45–68.

Mork, Knut A. 1989. "Oil and the Macroeconomy When Prices Go Up and Down: An Extension of Hamilton's Results." *Journal of Political Economy* 97 (3): 740–44.

——————————. 1994. "Business Cycles and the Oil Market." *Energy Journal* 15 (4): 15–38.