Economic Perspectives

2  Income inequality and redistribution in five countries

21  The expectations trap hypothesis

40  Subordinated debt as bank capital: A proposal for regulatory reform

54  Unemployment and wage growth: Recent cross-state evidence
<table>
<thead>
<tr>
<th>Page</th>
<th>Title</th>
<th>Authors</th>
<th>Abstract</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Income inequality and redistribution in five countries</td>
<td>Mariacristina De Nardi, Liqian Ren, and Chao Wei</td>
<td>This article studies income inequality in five countries and compares the redistributive consequences of taxes and transfers across these countries.</td>
</tr>
<tr>
<td>21</td>
<td>The expectations trap hypothesis</td>
<td>Lawrence J. Christiano and Christopher Gust</td>
<td>This article explores a hypothesis about the take-off in inflation in the early 1970s. According to the expectations trap hypothesis, the Fed was driven to high money growth by a fear of violating the expectations of high inflation that existed at the time. The authors argue that this hypothesis is more compelling than the Phillips curve hypothesis, according to which the Fed produced the high inflation as an unfortunate byproduct of a conscious decision to jump start a weak economy.</td>
</tr>
<tr>
<td>40</td>
<td>Subordinated debt as bank capital: A proposal for regulatory reform</td>
<td>Douglas D. Evanoff and Larry D. Wall</td>
<td>Industry observers have proposed increasing the role of subordinated debt in bank capital requirements as a means to increase market discipline. A recent Federal Reserve System Task Force evaluated the characteristics of such proposals. Here, the authors take the next step and offer a specific sub-debt proposal. They describe how it would operate and what changes it would require in the regulatory framework.</td>
</tr>
<tr>
<td>54</td>
<td>Unemployment and wage growth: Recent cross-state evidence</td>
<td>Daniel Aaronson and Daniel Sullivan</td>
<td>This article shows that even in recent years there is a relatively robust, negative cross-state correlation between appropriate measures of unemployment and wage growth.</td>
</tr>
</tbody>
</table>

Audio tapes for 2000 Bank Structure Conference
Introduction and summary

Policymakers designing or changing a country’s tax and transfer system aim at redistributing income and supporting the living standards of low-income families, while at the same time encouraging work effort and economic self-sufficiency. Indeed, there is a tradeoff between redistribution and efficiency: Economic theory suggests that transferring more income to the poor tends both to reduce their work effort and to distort the economic decisions of those who are taxed to provide the revenues that are being redistributed. There are several reasons why a government might want to redistribute income. Some of these are linked to the fact that people face different opportunities and different outcomes.

The government might want to provide insurance to its citizens against different outcomes, for example, sickness or unemployment, because in some cases private markets cannot work well. Moreover, not everybody enjoys the same opportunities in life; for example, people from poor family backgrounds are at a disadvantage relative to those from wealthier backgrounds, and transfers are a way to partly offset these differences.¹

For historical and social reasons, different countries put different weights on the costs and benefits of redistributing income. Traditionally, Anglo-Saxon countries have a relatively low degree of government intervention in the economy and place more emphasis on incentives, while in many European countries, we see relatively more government redistribution, greater provision of public goods, and more emphasis on equality of opportunities and outcomes. Our goal in this article is to look at different countries, study their redistribution policies, and discuss the effects of the redistribution/incentives tradeoff. Since we want to look at countries that display different degrees of government intervention, we pick countries belonging to both traditions. We focus on a small number of countries to study these issues in detail: the U.S., Canada, Germany, Sweden, and Finland. Our country choices are also limited by the availability of comparable data.

The link between the distribution of income and taxes and transfers is a complex one. Households in each country decide how hard to work, when to retire, and how much to consume and save, taking into account the incentives and disincentives provided by the structure of taxes and transfers in their country. Therefore, the distribution of labor income is itself endogenous and the actual measure of taxes and transfers depends on the labor and saving decisions of the households. Moreover, the distribution of labor income depends on the distribution of human capital, and the government, for example, by subsiding education, can have an impact on it.²

We focus on distribution of income across working-age households in these five countries because we are interested in labor income (earnings) inequality, abstracting from normal retirement decisions. In fact, at some age most people are retired and their labor income drops while their gross income is supplemented by social security payments, pensions, and other income sources. Looking only at households of working age, however, we ignore another important aspect of redistribution: social security transfers to older people.

¹ Mariacristina De Nardi is an economist at the Federal Reserve Bank of Chicago. Liqian Ren is an associate economist at the Federal Reserve Bank of Chicago. Chao Wei is a Ph.D. student at Stanford University. The authors would like to thank Marco Bassetto, Marco Cagetti, and David Marshall for helpful comments and Paul Alkemade and Dennis Sullivan for help with the dataset.
We study income inequality in these five countries and use different income measures to compare the redistributive consequences of taxes and transfers. We also discuss their likely effects on the households' labor, early retirement, and savings decisions. The distinction between transfers and taxes is interesting because transfers are typically not just connected to income, but may be means tested (both asset and income based) or based on a specific condition (for example, being unemployed or a single parent). Taxes are typically not related to means testing and depend much less on specific conditions. They rely mostly on income as the screening signal. Different mixes of taxes and transfers thus correspond to different screening mechanisms employed by each country in redistributing resources and, possibly, different redistributive goals.

All of the measures of income we look at are unequally distributed across countries and their distributions are concentrated and skewed. The U.S. displays the most unequal labor income distribution among the five countries, followed by Finland, Canada, Sweden, and Germany in that order. As we mentioned above, the distribution of labor income depends on the tax and transfer system, as well as on the distribution of human capital. Human capital is linked to education, which in turn is influenced by government subsidies. It is interesting to see that, as a result of all of these forces, the distribution of labor earnings in the countries that traditionally have been more concerned with redistribution (Finland and Sweden) is not necessarily more equal than it is in countries that belong to the Anglo-Saxon tradition of low government intervention (the U.S. and perhaps Canada). Finland is one obvious example of a country with high government intervention and high labor income inequality. Our research indicates that this is partly due to a more pronounced pattern of early retirement in Finland than in all of the other countries. Also, economic theory suggests that unemployment benefits discourage job search and work effort. This could translate into a larger number of unemployed or underemployed, which increases measured inequality in labor earnings.

Even after taxes and transfers, the U.S. displays by far the most unequal distribution for disposable income, followed by Canada, Germany, Finland, and Sweden. According to our data, and consistent with the distinction we discussed above, Finland reduces labor income inequality the most, followed by Sweden, Canada, Germany, and the U.S. Interestingly, Germany engages in little redistribution, but has the most equal distribution of labor earnings among these countries.

Not only do governments redistribute income differently, but they also use different instruments. In order to reduce labor income inequality, Finland and Sweden rely on a very progressive transfer system, while their tax system turns out to be very close to proportional (that is, close to a flat tax rate regime). At the opposite extreme, the U.S. uses taxes and transfers with approximately the same degree of progressivity. Canada and Germany are somewhere in between these extremes, with Canada relying more heavily on progressive transfers than Germany.

The progressivity of the tax and transfer systems is an important indicator of the resulting distortions in households' economic decisions. Another important indicator is given by the total amount of resources redistributed by the government in each country. As a measure, we can use the income tax faced by the average working age household. In our samples the average income tax rates are 16 percent in the U.S., 17 percent in Germany, 21 percent in Canada, 23 percent in Finland, and 25 percent in Sweden. In this sample, the countries with higher average income tax (Finland and Sweden) are also the ones with the least progressive tax systems. Government transfers (social insurance plus means-tested) as a fraction of gross income for the average working age household provide the same ordering of magnitude for redistribution as the average income tax. The average fractions of government transfers are 3 percent in the U.S., 6 percent in Germany, 8 percent in Canada, 15 percent in Finland, and 19 percent in Sweden.

We also look at the impact of transfers, conditional on the labor earnings level. For those in the bottom 10 percent of the labor earnings distribution in the U.S. and Canada, means-tested transfers, rather than social insurance transfers, are the main source of gross income. In contrast, in the other countries, and especially in Sweden, the main source of gross income for the poorest segment of the population is in the form of social insurance transfers.

Looking at the structure of earnings and transfers over the life cycle within each country, we find evidence that Finland and Sweden provide stronger incentives toward early retirement because of both social security and the structure of pension schemes. This explains some of the inequality we observe in the labor earnings distribution in these two countries; once people retire, their labor earnings drop. At the opposite extreme, our data suggest that there is less incentive to retire early in Germany and the U.S.

Our findings are thus consistent with the prediction from economic theory that greater redistribution through taxes and transfers is achieved at the cost of
greater distortions on labor supply and early retirement decisions. Consistent with other theoretical work, we also find that high redistribution countries rely heavily on instruments other than income taxes, such as transfers based on special conditions or means testing, to achieve high levels of redistribution while keeping distortions as low as possible for the beneficiaries. This, however, is costly because it generates the need to monitor eligibility. For example, Sweden has special agencies that monitor the job search efforts of the unemployed.

Germany is an interesting case. The level of redistribution through taxes and transfers is low. However, the distribution of labor earnings in Germany is remarkably more equal than in the other countries we consider here. Evidently the government is using other instruments to achieve this level of equality, possibly more equal access to public education. Another reason the distribution of earnings may be more equal is the presence of powerful unions, which typically favor a flat wage structure that enhances security at the expense of incentives.

**Definitions of income**

In this section we review the different definitions of income we use throughout the article and the information they convey. Our unit of analysis is the household, and the first measure of income we consider is labor income (earnings). This includes gross wage, salary income, and farm and nonfarm self-employment income. This measure provides us with information on the outcome of labor supply and early retirement decisions. Observing a large number of households with little or no earnings is an indication of high unemployment and/or a low participation rate. High levels of concentration in earnings might reflect a more unequal distribution of human capital and education in the population.

Our second measure of income is factor income which, besides earnings, includes cash property income (that is, cash interest, rents, dividends, and annuities) and royalties, but excludes capital gains and all other forms of lump-sum payments. Factor income, including income from capital, gives us a more comprehensive measure of income and provides indirect information on people’s assets and, hence, saving decisions.

Another measure of income is gross income, which adds social and private transfers to factor income. Government transfers might be an important channel through which the government redistributes income. Comparing the distribution of factor income with the one for gross income, we can study the effects of government transfers across different countries. Finally, we calculate disposable income by subtracting income taxes, mandatory employee contributions, and mandatory contributions for the self-employed from gross income. Disposable personal income provides a measure of the resources that households can actually allocate to either savings or consumption after taxes are paid and allows us to compare the progressivity of tax systems across different countries.

All of our statistics are based on total family income, without correcting for the number of family members. We also performed the computations taking into account family size to check whether different demographic patterns across countries affect our conclusions. To do so, we followed the “equivalence scale” literature and divided total family income for each family by the total number of family components, raised to the power $\alpha$. This method is meant to take into account that economies of scale arise as the size of the household increases. Our conclusions were not affected by this transformation.

**The data**

We use the Luxembourg Income Study (LIS) dataset. LIS collects existing household income surveys data from 25 countries and makes them comparable as much as possible in terms of data definition. The LIS dataset for the U.S. is based on the March Current Population Survey (CPS), the one for Canada on the Survey of Consumer Finances, the one for Germany on the German Socio-Economic Panel Study, the one for Sweden on the Income Distribution Survey, and the one for Finland on the Income Distribution Survey. The LIS provides data in waves; most of the datasets we use belong to the fourth wave. We use 1994 data for the U.S., Canada, and Germany and 1995 data for Finland. We use 1992 data for Sweden, because the 1995 Swedish dataset is still under revision.

The dataset has some limitations. These mainly stem from the fact that the data for the various countries come from existing datasets and might differ in the questions asked, their design, the definition of the household, and other important dimensions. While LIS aims at harmonizing the data so that the effect of these discrepancies is reduced, some differences will persist. Our minimum requirement to include a country was to have data on gross earnings, transfers, and taxes. This criterion alone excluded many countries, such as Italy and France, for which the only data available are net of taxes.

We provide a technical description of the country-specific datasets and their construction in the appendix.
LIS does not provide this information for the specific waves we use. We still report it, indicating to which year it refers, since it provides insight on the quality of the data across countries.

**An overview of income inequality across countries**

As we said earlier, we are interested in labor income inequality and redistribution. We do not have data on retirement status for all countries. Therefore, we concentrate on households whose head is of working age (25 to 60 years old, table 1). To study the possible effects of different patterns of early retirement on the income distribution, we also look at the subset of families whose head is 25 to 50 years of age (table 2). This will make quite a difference in the income distribution of some of the countries we consider but it will not matter much for others. We provide evidence in a later section that this is, indeed, related to early retirement decisions.

Tables 1 and 2 show that for both subsamples, earnings, factor income, gross income, and disposable income are unequally distributed across households in all of the countries and their distributions are concentrated and skewed (there are a large number of people with little and a small number of people with really large income of any type). The tables also show that governments redistribute with different strength and using different instruments.

The first column of each table reports the fraction of people with zero or negative earnings, factor income, gross income, and disposable income. In the dataset, all of the people with negative earnings are households with self-employment income in financial trouble.  

Looking at table 1 we see that the fraction of households at zero or negative earnings varies somewhat across these countries, with Finland having the highest fraction (9.7 percent) and Germany the lowest (7.0 percent). However, once all sources of income are taken into account and taxes are subtracted, this fraction drops below 1 percent for all countries, with the U.S. having the highest fraction of households with zero or negative disposable income (.9 percent) and Finland the lowest (.1 percent). Comparing the number of people with zero or negative earnings and factor income, we see that in all countries the fraction of people in this category falls when cash property income is added.  

<table>
<thead>
<tr>
<th>Country and variable</th>
<th>Fraction with zero or negative</th>
<th>Concentration Gini</th>
<th>p80/p20</th>
<th>Percentile location of mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>United States</td>
<td>Earnings 7.7</td>
<td>0.46</td>
<td>23</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td>Factor income 6.1</td>
<td>0.46</td>
<td>23</td>
<td>61</td>
</tr>
<tr>
<td></td>
<td>Gross income 0.9</td>
<td>0.42</td>
<td>12</td>
<td>62</td>
</tr>
<tr>
<td></td>
<td>Disposable income 0.9</td>
<td>0.39</td>
<td>9</td>
<td>60</td>
</tr>
<tr>
<td>Canada</td>
<td>Earnings 8.9</td>
<td>0.42</td>
<td>24</td>
<td>56</td>
</tr>
<tr>
<td></td>
<td>Factor income 7.7</td>
<td>0.42</td>
<td>22</td>
<td>56</td>
</tr>
<tr>
<td></td>
<td>Gross income 0.2</td>
<td>0.35</td>
<td>8</td>
<td>58</td>
</tr>
<tr>
<td></td>
<td>Disposable income 0.2</td>
<td>0.32</td>
<td>6</td>
<td>56</td>
</tr>
<tr>
<td>Germany</td>
<td>Earnings 7.0</td>
<td>0.38</td>
<td>13</td>
<td>56</td>
</tr>
<tr>
<td></td>
<td>Factor income 6.2</td>
<td>0.39</td>
<td>14</td>
<td>57</td>
</tr>
<tr>
<td></td>
<td>Gross income 0.2</td>
<td>0.34</td>
<td>7</td>
<td>59</td>
</tr>
<tr>
<td></td>
<td>Disposable income 0.2</td>
<td>0.30</td>
<td>5</td>
<td>58</td>
</tr>
<tr>
<td>Sweden</td>
<td>Earnings 7.6</td>
<td>0.39</td>
<td>19</td>
<td>56</td>
</tr>
<tr>
<td></td>
<td>Factor income 3.7</td>
<td>0.39</td>
<td>17</td>
<td>57</td>
</tr>
<tr>
<td></td>
<td>Gross income 0.3</td>
<td>0.29</td>
<td>5</td>
<td>54</td>
</tr>
<tr>
<td></td>
<td>Disposable income 0.3</td>
<td>0.27</td>
<td>4</td>
<td>53</td>
</tr>
<tr>
<td>Finland</td>
<td>Earnings 9.7</td>
<td>0.43</td>
<td>39</td>
<td>56</td>
</tr>
<tr>
<td></td>
<td>Factor income 7.8</td>
<td>0.44</td>
<td>36</td>
<td>57</td>
</tr>
<tr>
<td></td>
<td>Gross income 0.0</td>
<td>0.32</td>
<td>6</td>
<td>57</td>
</tr>
<tr>
<td></td>
<td>Disposable income 0.1</td>
<td>0.29</td>
<td>5</td>
<td>55</td>
</tr>
</tbody>
</table>

Notes: The Gini coefficient is a measure of inequality which varies between 0 and 1. 0 indicates perfect equality, 1 indicates perfect inequality (see box 1). The variable p80/p20 is a measure of social distance. It measures the ratio of the average income of the richest and poorest 20 percent of the population.

The second column reports the Gini coefficient (see box 1), which is a measure of inequality. The U.S. displays the highest concentration for all income measures, Germany has the least concentrated earnings distribution, and Sweden has the least concentration in the gross and disposable income distributions. There is some evidence that Germany achieves redistribution using some other mechanism that makes labor earnings more equal.

The drop in the Gini index from one row to the next measures the reduction in inequality. We see that Finland achieves more redistribution (its Gini coefficient for disposable income is 34 percent lower than its Gini coefficient for factor income), most of which comes from transfers. Sweden is quite close to Finland, both in the size of the redistribution and the use of transfers to achieve it. At the opposite extreme, in the U.S. the combined effect of taxes and transfers reduces the factor income Gini coefficient by 15 percent, and transfers cause only about half of the reduction. Canada and Germany are somewhere in between, with Canada relying more heavily on transfers than Germany.

The fourth column of the tables reports another measure of concentration. Let us take earnings: p80/p20 is the ratio between the total earnings of the richest 20 percent, divided by the total earnings of the poorest 20 percent. This is a measure of “social distance,” comparing the richest population segment with the poorest.

In table 1, the p80/p20 earnings ratio varies between a high of 39 for Finland and a low of 13 for Germany. The ratio in Finland is high not because the richest people make more here than in the other countries, but because the average earnings of the poorest 20 percent are low compared with the other countries. After taxes and transfers, the p80/p20 ratio for disposable income falls noticeably. In all countries but the U.S. this is mostly due to transfer systems that increase significantly the gross income of the poorest, rather than to tax systems that reduce more than proportionally the average disposable income of the richest. The p80/p20 for disposable income is highest in the U.S. (9) and lowest in Sweden (4).

Comparing table 1 and table 2, we see that restricting our sample to households whose head is 50 and younger makes a difference, especially for Finland, Canada, and Sweden. For example, p80/p20, the measure of social distance from richest 20 percent to poorest 20 percent, drops from 39 to 21 for Finland when we change the upper age limit from 60 to 50. However, it makes little difference for the U.S. and no difference for Germany. This suggests that people might retire earlier in some countries than in others. According to the Gini coefficient for earnings reported in table 2, the U.S. is still the country with the highest earnings inequality, followed by Canada, Finland, Sweden, and Germany.

The last column, percentile location of mean, provides information on the skewness of the distribution.

### TABLE 2

<table>
<thead>
<tr>
<th>Country and variable</th>
<th>Fraction with zero or negative Gini</th>
<th>Concentration p80/p20</th>
<th>Percentile location of mean</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>United States</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Earnings</td>
<td>6.8</td>
<td>0.45</td>
<td>21</td>
</tr>
<tr>
<td>Factor income</td>
<td>5.8</td>
<td>0.45</td>
<td>21</td>
</tr>
<tr>
<td>Gross income</td>
<td>0.9</td>
<td>0.42</td>
<td>11</td>
</tr>
<tr>
<td>Disposable income</td>
<td>0.9</td>
<td>0.38</td>
<td>9</td>
</tr>
<tr>
<td><strong>Canada</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Earnings</td>
<td>7.6</td>
<td>0.41</td>
<td>19</td>
</tr>
<tr>
<td>Factor income</td>
<td>7.1</td>
<td>0.40</td>
<td>18</td>
</tr>
<tr>
<td>Gross income</td>
<td>0.2</td>
<td>0.34</td>
<td>7</td>
</tr>
<tr>
<td>Disposable income</td>
<td>0.2</td>
<td>0.31</td>
<td>6</td>
</tr>
<tr>
<td><strong>Germany</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Earnings</td>
<td>5.9</td>
<td>0.38</td>
<td>12</td>
</tr>
<tr>
<td>Factor income</td>
<td>5.4</td>
<td>0.38</td>
<td>12</td>
</tr>
<tr>
<td>Gross income</td>
<td>0.0</td>
<td>0.34</td>
<td>6</td>
</tr>
<tr>
<td>Disposable income</td>
<td>0.0</td>
<td>0.30</td>
<td>5</td>
</tr>
<tr>
<td><strong>Sweden</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Earnings</td>
<td>6.7</td>
<td>0.39</td>
<td>17</td>
</tr>
<tr>
<td>Factor income</td>
<td>3.5</td>
<td>0.39</td>
<td>16</td>
</tr>
<tr>
<td>Gross income</td>
<td>0.3</td>
<td>0.29</td>
<td>4</td>
</tr>
<tr>
<td>Disposable income</td>
<td>0.3</td>
<td>0.27</td>
<td>4</td>
</tr>
<tr>
<td><strong>Finland</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Earnings</td>
<td>7.2</td>
<td>0.40</td>
<td>21</td>
</tr>
<tr>
<td>Factor income</td>
<td>6.3</td>
<td>0.41</td>
<td>20</td>
</tr>
<tr>
<td>Gross income</td>
<td>0.0</td>
<td>0.31</td>
<td>5</td>
</tr>
<tr>
<td>Disposable income</td>
<td>0.1</td>
<td>0.28</td>
<td>4</td>
</tr>
</tbody>
</table>

Notes: The Gini coefficient is a measure of inequality which varies between 0 and 1. 0 indicates perfect equality, 1 indicates perfect inequality (see box 1). The variable p80/p20 is a measure of social distance. It measures the ratio of the average income of the richest and poorest 20 percent of the population.

This measure reveals that in the U.S. the distributions are more skewed, both before and after taxes and transfers. The distributions of earnings and factor income are similarly skewed in Canada, Germany, Sweden, and Finland, while Sweden displays less skewness in its distribution of disposable income.

**Using Lorenz curves to better understand inequality**

Figure 1 compares the Lorenz curve for earnings across the five countries. As we explain in box 1, the Lorenz curve provides more information than the Gini index, which is a summary measure of inequality. It is

---

**BOX 1**

**Lorenz curve and Gini coefficient**

The Lorenz curve provides information on inequality. To draw it, we first sort the households by their income, starting with the ones with the lowest income. We then plot the relationship between the cumulative percentage of the population (on the horizontal axis) and the proportion of total income earned by each cumulative percentage (on the vertical axis). Figures a and b show the Lorenz curve for the two extreme cases of perfect equality and highest inequality. In the case of perfect equality everybody earns the same proportion of total income, and the Lorenz curve coincides with the 45-degree line (see figure a). In the case of perfect inequality, just one family earns all of the total income in the economy. All households except the last one earn no income, and hence the cumulative proportion of income earned stays at zero. The Lorenz curve stays flat until the very last household is reached; then it jumps to 100, since the last family earns all of the income in the economy.

In real life we observe intermediate cases, in which some households earn more and others less, and the Lorenz curve lies between the perfect equality and the perfect inequality lines (figure c).

The Gini coefficient is a summary statistic of inequality derived from the Lorenz curve. It is defined as the ratio of area A (see figure c: the area between the Lorenz curve and the perfect equality line) to area A + B (the area between the perfect equality and perfect inequality lines). The Gini coefficient varies between zero and one; it is equal to zero in the case of perfect equality (every household earns the same) and equal to one in the case of perfect inequality (one household earns everything). Therefore, the Gini coefficient provides a summary measure of inequality over the whole range of the distribution.
interesting to observe not only the ordering of the curves for the various countries (the ones that lie to the right are the farthest from the 45-degree line and thus indicate a country with more inequality), but also whether the lines cross and where. Until the thirty-fifth percentile, Finland is the country in which the poorest families earn the smallest fraction of total earnings. From that percentile on, the U.S. emerges as having greater income inequality than Finland or any of the other countries we study.

Economic theory (for a survey, see Mortensen and Pissarides, 1999) suggests that workers’ labor decisions depend, among other things, on the social security safety net that is in place: In countries with more generous social insurance systems (such as unemployment benefits), workers will be pickier and there will be more people with zero earnings, since they receive transfers from the government. In this case, the workers are deciding not to work, or not to work for a longer period because of the availability of benefits; thus, they may be better off than the workers in countries that do not offer such generous benefits. The incentives to retire early also affect the number of people at low levels of earnings. These incentives differ across countries, and we provide evidence that they are particularly strong in Finland.

Looking at the earnings of households between the fortieth and eightieth percentiles, the ordering of the countries from most equal to most unequal is Germany, Sweden, Canada, Finland, and the U.S.

Figure 2 displays the Lorenz curves for gross income across the five countries. After adding private and government transfers, the U.S. displays the most concentrated distribution by far for all percentiles. Until the eighty-fifth percentile, the ordering of gross income inequality from the most equal to the most unequal is Sweden, Finland, Germany, Canada, and the U.S. After adding transfers, the poorest people in the other countries are noticeably better off than in the U.S. This is not the case for the earnings distributions in figure 1. As we discussed for table 1, transfers go a long way in redistributing income, especially at the lower levels of earnings. For all countries but the U.S. and Germany, they are the instrument most used to redistribute income. However, economic theory predicts that a
generous transfer system influences labor supply and early retirement decisions, increasing the number of people at zero earnings and reducing labor supply even at higher levels.

Figure 3 shows the Lorenz curves for disposable income. As in figure 2, the Lorenz curve for the U.S. is by far the most concentrated at all percentiles. The Lorenz curves for Sweden, Finland, and Germany are closer than the ones for gross earnings and almost coincide for the poorest 60 percent of the population. High redistribution countries rely heavily on instruments other than income taxes, such as transfers based on special conditions or means testing, to achieve high levels of redistribution while keeping distortions as low as possible for the beneficiaries.

As we mentioned earlier, however, this is costly because it generates the need to monitor eligibility.

Figures 4 to 8 display the Lorenz curves for earnings, gross income, and disposable income within each country. Comparing the figures, we see that the U.S. and Germany redistribute income across households using transfers and taxes roughly with the same intensity, with transfers having the strongest impact for families below the median earner family and taxes becoming more redistributive for families above the twenty-fifth percentile. In Canada, the effect of transfers shifts the Lorenz curve for gross income more than it does in the U.S. Both Sweden and Finland have very high levels of redistributions by means of transfers, also for families high up in the distribution, while taxation shifts the Lorenz curve relatively little in both cases. We should notice that proportional taxation (income is taxed at the same marginal rate, regardless of the income level) and proportional transfers do not shift the Lorenz curve and do not change the Gini coefficient. Conversely, progressive taxation (higher income is
taxed at a higher marginal rate) and transfers do. Therefore, our comparison shows that the Swedish and Finnish tax systems are effectively close to a proportional tax and all of the progressivity is achieved through transfers. Taxation is more progressive in the U.S., Canada, and Germany.

So far, we have discussed the progressivity of the tax and transfer systems in our five countries based on how they change the relative position of the households in the income distribution. However, this criterion does not give us much information about the magnitude of the income that changes hands in the economy. From the last columns of tables 3 to 7, we can look at another measure of redistribution within each country: aggregate taxes and transfers as a fraction of aggregate gross income. Looking at this criterion, we see that total transfers are 6 percent of gross income of the working age families in the U.S., 11 percent in Canada, 7 percent in Germany, 19 percent in Sweden, and 21 percent in Finland. For income taxes, the numbers are 16 percent, 21 percent, 17 percent, 25 percent, and 23 percent of gross income, respectively. The magnitude of these flows provides the same ordering of strength of redistribution across countries suggested by the Lorenz curves and the Gini coefficients.

**Labor earnings and redistribution**

Tables 3 to 7 provide more detail on earnings, taxes, and transfers for households whose head is 25 to 60 years of age, conditional on labor earnings quartiles. In each table, the columns provide information about a number of households, classified according to their relative position in the earnings distribution of the total sample of households: the poorest 10 percent, the quartiles, the richest 10 percent, and the population as a whole. We study average earnings, gross income, and disposable income for the households in each category. To better understand how redistribution takes place within the quartiles of the earnings distribution, we also analyze the sources of disposable income and tax payments.

We distinguish among various income sources. The first three are gross wage and salary income.
(labor), income from self-employment (business), and cash property income. We then distinguish among several transfer components. Social insurance transfers include sick, accident, and disability pay, social retirement benefits (even if the household head is of working age, he or she may go into early retirement or another family member might receive such payments), child or family allowances, unemployment compensation, maternity pay, military/veteran/war benefits, and other social insurance. Means-tested transfers include both cash and near-cash benefits. Pensions include private pensions and public sector pensions. Private pensions are employer payments for retirement that may supplement social security transfers. Self-employment pension plans are included, if they are designed to supplement social security, for example, individual retirement accounts (IRAs). Public sector pensions include pensions for public employees and do not include amounts coming from social security benefits for the aged or survivors. Private transfers include alimony or child support and other regular private income.

We then report income taxes. We do not have information on employee and self-employed contributions for all five countries. The comparison between income tax rates is likely to carry over to the entire tax system, as the income tax is the most progressive component of the tax code.

We also report some demographic characteristics of households in the different earnings quartiles.

### The U.S.

As table 3 shows, the average household at the bottom 10 percent of the earnings distribution in the U.S. earns $275 from labor income, which amounts to a disposable income of $9,090 after taxes and transfers. Less than 3 percent of the household’s gross income comes from earnings, while 86.4 percent derives from transfers. For this group, means-tested transfers account for the largest share of transfers (37.4 percent), followed by social insurance (26.3 percent) and pensions (13.4 percent). Consistent with the observation that lifetime earnings follow an inverted U shape, the 10 percent of households with the lowest earnings include a disproportionate share of the youngest and oldest population segments. Young people are still accumulating human capital and trying to climb up the earnings distribution. The relatively high fraction

<table>
<thead>
<tr>
<th>Household characteristics</th>
<th>Bottom 10%</th>
<th>1st</th>
<th>2nd</th>
<th>3rd</th>
<th>4th</th>
<th>Top 10%</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Earnings, in dollars</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average earnings</td>
<td>275</td>
<td>6,009</td>
<td>24,494</td>
<td>43,415</td>
<td>89,184</td>
<td>122,085</td>
<td>40,676</td>
</tr>
<tr>
<td>Average gross income</td>
<td>9,448</td>
<td>12,295</td>
<td>27,220</td>
<td>46,352</td>
<td>94,395</td>
<td>129,545</td>
<td>44,965</td>
</tr>
<tr>
<td>Average disposable income</td>
<td>9,090</td>
<td>11,320</td>
<td>22,791</td>
<td>37,016</td>
<td>68,258</td>
<td>89,275</td>
<td>34,773</td>
</tr>
<tr>
<td>Sources of gross income, %</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Labor</td>
<td>2.7</td>
<td>45.2</td>
<td>84.5</td>
<td>88.4</td>
<td>87.6</td>
<td>85.9</td>
<td>84.4</td>
</tr>
<tr>
<td>Business</td>
<td>0.2</td>
<td>3.6</td>
<td>5.5</td>
<td>5.3</td>
<td>6.9</td>
<td>8.4</td>
<td>6.0</td>
</tr>
<tr>
<td>Cash property income</td>
<td>10.7</td>
<td>6.1</td>
<td>2.5</td>
<td>2.6</td>
<td>3.9</td>
<td>4.6</td>
<td>3.5</td>
</tr>
<tr>
<td>Total transfers</td>
<td>86.4</td>
<td>45.1</td>
<td>7.6</td>
<td>3.7</td>
<td>1.6</td>
<td>1.2</td>
<td>6.0</td>
</tr>
<tr>
<td>Social insurance</td>
<td>26.3</td>
<td>16.5</td>
<td>3.5</td>
<td>1.6</td>
<td>0.7</td>
<td>0.4</td>
<td>2.4</td>
</tr>
<tr>
<td>Means-tested</td>
<td>37.4</td>
<td>15.9</td>
<td>1.0</td>
<td>0.2</td>
<td>0.0</td>
<td>0.0</td>
<td>1.3</td>
</tr>
<tr>
<td>Pensions</td>
<td>13.4</td>
<td>7.1</td>
<td>1.7</td>
<td>1.2</td>
<td>0.7</td>
<td>0.6</td>
<td>1.4</td>
</tr>
<tr>
<td>Private</td>
<td>7.3</td>
<td>4.8</td>
<td>1.2</td>
<td>0.6</td>
<td>0.2</td>
<td>0.1</td>
<td>0.8</td>
</tr>
<tr>
<td>Income tax, %</td>
<td>3.5</td>
<td>4.0</td>
<td>9.4</td>
<td>13.0</td>
<td>21.3</td>
<td>25.2</td>
<td>16.2</td>
</tr>
<tr>
<td>Average number of earners</td>
<td>0.3</td>
<td>0.8</td>
<td>1.4</td>
<td>1.8</td>
<td>2.2</td>
<td>2.2</td>
<td>1.5</td>
</tr>
<tr>
<td>Average household size</td>
<td>2.3</td>
<td>2.3</td>
<td>2.4</td>
<td>2.9</td>
<td>3.3</td>
<td>3.3</td>
<td>2.7</td>
</tr>
<tr>
<td>Age of household head, %</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25–34</td>
<td>32.6</td>
<td>37.9</td>
<td>38.2</td>
<td>29.4</td>
<td>17.1</td>
<td>13.3</td>
<td>30.7</td>
</tr>
<tr>
<td>35–49</td>
<td>39.2</td>
<td>39.8</td>
<td>43.5</td>
<td>49.7</td>
<td>56.9</td>
<td>57.4</td>
<td>47.5</td>
</tr>
<tr>
<td>50–60</td>
<td>28.2</td>
<td>22.3</td>
<td>18.2</td>
<td>20.9</td>
<td>26.0</td>
<td>29.3</td>
<td>21.8</td>
</tr>
<tr>
<td>Average age, years</td>
<td>41.4</td>
<td>39.8</td>
<td>39.3</td>
<td>40.9</td>
<td>43.4</td>
<td>44.4</td>
<td>40.8</td>
</tr>
</tbody>
</table>

of older people (50 to 60 years old) among the lowest earners suggests that a significant number of people in our sample are taking early retirement. As we mentioned earlier, this is a common feature across countries, although it is much more common in Finland and Sweden.

Looking at the overall distribution, we see that transfers decline quickly as earnings increase, with means-tested transfers declining even more quickly. The share of pensions also declines throughout the distribution.

The structure of taxation is very progressive, with the average tax rate going from 3.5 percent for the poorest 10 percent, to 25 percent for the richest 10 percent. However, the average tax rate in the U.S. is low, compared with the other countries we look at.

**Canada**

Table 4 shows that Canada has a more generous transfer system than the U.S. Both social insurance and means-tested transfers are larger in Canada, but while social insurance transfers decline more slowly as earnings increase, means-tested ones do so more quickly, as households in the second quartile of both distributions receive less than 1 percent of their gross income from this source. The share of pension income across the distribution looks remarkably similar to the one in the U.S. even though in Canada the fraction of people between 50 and 60 years of age is larger.

The Canadian income tax regime is almost as progressive as the U.S. one. In particular, households at the top 10 percent of the distribution pay an average income tax of 28 percent, compared with 25 percent in the U.S., although for the whole population the average rate is 21 percent in Canada and 16 percent in the U.S.

**Germany**

The fraction of gross income coming from government transfers (social insurance plus means-tested) for the average household in the total population is 6.4 percent, compared with 3.7 percent in the U.S. and 8 percent in Canada. Interestingly, at the bottom 10 percent of the earnings distribution the share of transfers due to social insurance is larger than the means-tested share in Germany, unlike in the U.S. and Canada.

### TABLE 4

**Canadian households ranked by earnings**

<table>
<thead>
<tr>
<th>Household characteristics</th>
<th>Bottom 10%</th>
<th>1st</th>
<th>2nd</th>
<th>3rd</th>
<th>4th</th>
<th>Top 10%</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Earnings, in U.S. dollars</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average earnings</td>
<td>68</td>
<td>5,088</td>
<td>22,753</td>
<td>37,856</td>
<td>68,148</td>
<td>88,188</td>
<td>33,408</td>
</tr>
<tr>
<td>Average gross income</td>
<td>11,004</td>
<td>13,472</td>
<td>27,281</td>
<td>40,972</td>
<td>71,363</td>
<td>91,807</td>
<td>38,230</td>
</tr>
<tr>
<td>Average disposable income</td>
<td>10,412</td>
<td>12,449</td>
<td>22,982</td>
<td>32,570</td>
<td>53,112</td>
<td>66,075</td>
<td>30,246</td>
</tr>
<tr>
<td>Sources of gross income</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Labor</td>
<td>0.5</td>
<td>32.7</td>
<td>76.5</td>
<td>87.4</td>
<td>88.1</td>
<td>86.3</td>
<td>80.9</td>
</tr>
<tr>
<td>Business</td>
<td>0.1</td>
<td>5.1</td>
<td>6.9</td>
<td>5.0</td>
<td>7.4</td>
<td>9.8</td>
<td>6.5</td>
</tr>
<tr>
<td>Cash property income</td>
<td>6.4</td>
<td>4.7</td>
<td>2.0</td>
<td>1.4</td>
<td>1.7</td>
<td>2.0</td>
<td>1.9</td>
</tr>
<tr>
<td>Total transfers</td>
<td>93.0</td>
<td>57.6</td>
<td>14.6</td>
<td>6.2</td>
<td>2.8</td>
<td>2.0</td>
<td>10.7</td>
</tr>
<tr>
<td>Social insurance</td>
<td>30.6</td>
<td>26.5</td>
<td>10.1</td>
<td>4.4</td>
<td>1.7</td>
<td>1.0</td>
<td>6.1</td>
</tr>
<tr>
<td>Means-tested</td>
<td>41.3</td>
<td>18.4</td>
<td>0.8</td>
<td>0.3</td>
<td>0.1</td>
<td>0.1</td>
<td>1.9</td>
</tr>
<tr>
<td>Pensions</td>
<td>14.1</td>
<td>7.4</td>
<td>1.7</td>
<td>0.8</td>
<td>0.4</td>
<td>0.3</td>
<td>1.4</td>
</tr>
<tr>
<td>Private</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Income tax, %</td>
<td>5.4</td>
<td>7.6</td>
<td>15.8</td>
<td>20.5</td>
<td>25.6</td>
<td>28.0</td>
<td>20.9</td>
</tr>
<tr>
<td>Average number of earners</td>
<td>0.1</td>
<td>0.8</td>
<td>1.6</td>
<td>1.9</td>
<td>2.4</td>
<td>2.6</td>
<td>1.7</td>
</tr>
<tr>
<td>Average household size</td>
<td>2.1</td>
<td>2.2</td>
<td>2.6</td>
<td>3.0</td>
<td>3.4</td>
<td>3.5</td>
<td>2.8</td>
</tr>
<tr>
<td>Age of household head, %</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25–34</td>
<td>28.0</td>
<td>32.1</td>
<td>35.9</td>
<td>29.4</td>
<td>19.2</td>
<td>13.7</td>
<td>29.2</td>
</tr>
<tr>
<td>35–49</td>
<td>39.9</td>
<td>42.0</td>
<td>44.8</td>
<td>51.0</td>
<td>55.5</td>
<td>56.8</td>
<td>48.3</td>
</tr>
<tr>
<td>50–60</td>
<td>32.2</td>
<td>25.9</td>
<td>19.3</td>
<td>19.5</td>
<td>25.3</td>
<td>29.5</td>
<td>22.5</td>
</tr>
<tr>
<td>Average age, years</td>
<td>43.0</td>
<td>41.3</td>
<td>39.8</td>
<td>40.7</td>
<td>43.1</td>
<td>44.4</td>
<td>41.2</td>
</tr>
</tbody>
</table>

### TABLE 5

German households ranked by earnings

<table>
<thead>
<tr>
<th>Household characteristics</th>
<th>Bottom 10%</th>
<th>1st</th>
<th>2nd</th>
<th>3rd</th>
<th>4th</th>
<th>Top 10%</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Earnings, in U.S. dollars</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average earnings</td>
<td>867</td>
<td>9,174</td>
<td>30,275</td>
<td>45,496</td>
<td>80,412</td>
<td>106,831</td>
<td>41,333</td>
</tr>
<tr>
<td>Average gross income</td>
<td>14,926</td>
<td>18,247</td>
<td>33,344</td>
<td>48,394</td>
<td>84,407</td>
<td>113,440</td>
<td>46,092</td>
</tr>
<tr>
<td>Average disposable income</td>
<td>13,584</td>
<td>15,602</td>
<td>24,236</td>
<td>33,921</td>
<td>54,320</td>
<td>70,518</td>
<td>32,016</td>
</tr>
<tr>
<td>Sources of gross income, %</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Labor</td>
<td>5.6</td>
<td>44.6</td>
<td>86.9</td>
<td>87.4</td>
<td>85.5</td>
<td>80.6</td>
<td>82.2</td>
</tr>
<tr>
<td>Business</td>
<td>0.2</td>
<td>5.7</td>
<td>3.9</td>
<td>6.6</td>
<td>9.8</td>
<td>13.6</td>
<td>7.5</td>
</tr>
<tr>
<td>Cash property income</td>
<td>21.4</td>
<td>8.2</td>
<td>1.3</td>
<td>2.0</td>
<td>3.1</td>
<td>4.7</td>
<td>3.0</td>
</tr>
<tr>
<td>Total transfers</td>
<td>72.8</td>
<td>23.4</td>
<td>6.2</td>
<td>3.1</td>
<td>1.3</td>
<td>0.9</td>
<td>4.8</td>
</tr>
<tr>
<td>Social insurance</td>
<td>36.8</td>
<td>23.4</td>
<td>6.2</td>
<td>3.1</td>
<td>1.3</td>
<td>0.9</td>
<td>4.8</td>
</tr>
<tr>
<td>Means-tested</td>
<td>23.5</td>
<td>11.9</td>
<td>1.2</td>
<td>0.5</td>
<td>0.1</td>
<td>0.1</td>
<td>1.6</td>
</tr>
<tr>
<td>Pensions</td>
<td>3.3</td>
<td>1.9</td>
<td>0.2</td>
<td>0.3</td>
<td>0.0</td>
<td>0.0</td>
<td>0.3</td>
</tr>
<tr>
<td>Private</td>
<td>9.3</td>
<td>4.4</td>
<td>0.4</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.6</td>
</tr>
<tr>
<td>Income tax, %</td>
<td>8.7</td>
<td>6.4</td>
<td>11.5</td>
<td>14.6</td>
<td>22.9</td>
<td>27.4</td>
<td>17.0</td>
</tr>
<tr>
<td>Average number of earners</td>
<td>0.3</td>
<td>0.8</td>
<td>1.4</td>
<td>1.8</td>
<td>2.1</td>
<td>2.1</td>
<td>1.5</td>
</tr>
<tr>
<td>Average household size</td>
<td>1.9</td>
<td>2.1</td>
<td>2.5</td>
<td>2.9</td>
<td>3.1</td>
<td>3.2</td>
<td>2.6</td>
</tr>
<tr>
<td>Age of household head, %</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25–34</td>
<td>37.2</td>
<td>39.4</td>
<td>38.0</td>
<td>29.3</td>
<td>19.8</td>
<td>10.7</td>
<td>31.6</td>
</tr>
<tr>
<td>35–49</td>
<td>29.7</td>
<td>31.8</td>
<td>36.8</td>
<td>46.0</td>
<td>49.5</td>
<td>51.7</td>
<td>41.0</td>
</tr>
<tr>
<td>50–60</td>
<td>33.1</td>
<td>28.8</td>
<td>25.1</td>
<td>24.7</td>
<td>30.7</td>
<td>37.7</td>
<td>27.3</td>
</tr>
<tr>
<td>Average age, years</td>
<td>42.0</td>
<td>40.8</td>
<td>40.4</td>
<td>41.5</td>
<td>43.9</td>
<td>45.9</td>
<td>41.7</td>
</tr>
</tbody>
</table>

Source: Luxembourg Income Study, 1994, dataset for Germany, Differdange, Luxembourg; Centre for Population, Poverty, and Policy Studies.

The share of gross income due to pensions is low in Germany; for example, at the bottom 10 percent it is only 3.3 percent, compared with about 14 percent in both the U.S. and Canada, despite the fact that the share of people ages 50 to 60 is larger in Germany. This reflects the fact that the German social security system is much less redistributive than in the other countries (see Börsch-Supan and Schnabel, 1999), so the share of payments that goes to the poorest segment of the population is lower.

As we said before, Germany is the country with the second least generous transfer system after the U.S. It is also the country with the second lowest average income tax, 17 percent of total gross income, compared with 16 percent in the U.S., 21 percent in Canada, and much higher rates in Sweden and Finland. However, the bottom 10 percent of households pay more taxes in Germany (8.7 percent) than in the U.S. (3.5 percent) or Canada (5.4 percent).

**Sweden**

In Sweden, 19 percent of average household gross income is due to transfers, compared with 6 percent in the U.S., 7 percent in Germany, and 11 percent in Canada.\(^{15}\)

Comparing tables 6 and 3 we see that the Swedish households at the bottom 10 percent of the earnings distribution have $223 in average earnings, compared with $275 in the U.S., but end up with an average disposable income of $19,750, compared with $9,090 in the U.S. They thus receive 92 percent of their gross income from transfers, the majority of which is social assistance (this, however, includes public pensions in Sweden), while a much smaller fraction is means tested. Swedish social security transfers remain large as earnings increase: The households in the top quartile of the earnings distribution receive 5 percent of their gross income from government transfers.

Correspondingly, the average income tax for the whole population is also much larger (25 percent) than in the countries we have discussed so far. Its structure is not very progressive, starting from an average rate of 16 percent for the bottom 10 percent up to 31 percent for the richest 10 percent.
**Swedish households ranked by earnings**

<table>
<thead>
<tr>
<th>Household characteristics</th>
<th>Bottom 10%</th>
<th>1st</th>
<th>2nd</th>
<th>3rd</th>
<th>4th</th>
<th>Top 10%</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Earnings, in U.S. dollars</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average earnings</td>
<td>223</td>
<td>7,010</td>
<td>28,120</td>
<td>44,315</td>
<td>76,646</td>
<td>96,233</td>
<td>39,020</td>
</tr>
<tr>
<td>Average gross income</td>
<td>23,593</td>
<td>26,798</td>
<td>36,960</td>
<td>53,925</td>
<td>84,404</td>
<td>104,351</td>
<td>50,519</td>
</tr>
<tr>
<td>Average disposable income</td>
<td>19,750</td>
<td>21,806</td>
<td>28,178</td>
<td>40,821</td>
<td>60,203</td>
<td>71,928</td>
<td>37,750</td>
</tr>
<tr>
<td>Sources of gross income, %</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Labor</td>
<td>0.8</td>
<td>23.5</td>
<td>73.7</td>
<td>79.4</td>
<td>88.9</td>
<td>90.2</td>
<td>74.9</td>
</tr>
<tr>
<td>Business</td>
<td>0.1</td>
<td>2.7</td>
<td>2.4</td>
<td>2.8</td>
<td>1.9</td>
<td>2.1</td>
<td>2.3</td>
</tr>
<tr>
<td>Cash property income</td>
<td>4.2</td>
<td>3.8</td>
<td>2.7</td>
<td>3.5</td>
<td>3.9</td>
<td>3.9</td>
<td>3.6</td>
</tr>
<tr>
<td>Total transfers</td>
<td>94.8</td>
<td>70.0</td>
<td>21.2</td>
<td>14.3</td>
<td>5.3</td>
<td>3.9</td>
<td>19.2</td>
</tr>
<tr>
<td>Social insurance</td>
<td>73.1</td>
<td>55.0</td>
<td>18.0</td>
<td>13.0</td>
<td>5.0</td>
<td>3.7</td>
<td>16.2</td>
</tr>
<tr>
<td>Means-tested</td>
<td>8.6</td>
<td>12.6</td>
<td>2.0</td>
<td>0.9</td>
<td>0.1</td>
<td>0.1</td>
<td>2.3</td>
</tr>
<tr>
<td>Pensions</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Private</td>
<td>3.1</td>
<td>2.5</td>
<td>1.2</td>
<td>0.5</td>
<td>0.1</td>
<td>0.1</td>
<td>0.7</td>
</tr>
<tr>
<td>Income tax, %</td>
<td>16.3</td>
<td>18.6</td>
<td>23.8</td>
<td>24.3</td>
<td>28.7</td>
<td>31.1</td>
<td>25.3</td>
</tr>
<tr>
<td>Average number of earners</td>
<td>0.5</td>
<td>1.0</td>
<td>1.3</td>
<td>1.7</td>
<td>2.1</td>
<td>2.1</td>
<td>1.5</td>
</tr>
<tr>
<td>Average household size</td>
<td>1.7</td>
<td>1.8</td>
<td>1.8</td>
<td>2.5</td>
<td>2.9</td>
<td>2.9</td>
<td>2.2</td>
</tr>
<tr>
<td>Age of household head, %</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25–34</td>
<td>31.1</td>
<td>39.5</td>
<td>39.9</td>
<td>30.7</td>
<td>13.0</td>
<td>8.2</td>
<td>30.8</td>
</tr>
<tr>
<td>35–49</td>
<td>37.9</td>
<td>36.9</td>
<td>40.1</td>
<td>47.6</td>
<td>55.9</td>
<td>59.5</td>
<td>45.1</td>
</tr>
<tr>
<td>50–60</td>
<td>31.0</td>
<td>23.5</td>
<td>20.0</td>
<td>21.7</td>
<td>31.0</td>
<td>32.3</td>
<td>24.1</td>
</tr>
<tr>
<td>Average age, years</td>
<td>41.6</td>
<td>39.7</td>
<td>39.4</td>
<td>40.9</td>
<td>44.7</td>
<td>45.8</td>
<td>41.2</td>
</tr>
</tbody>
</table>

Note: N/A indicates not available.

**Finland**

As we see from table 7, in Finland as in Sweden, the amount of transfer income is substantial and the part due to social insurance is generous throughout the earnings distribution. In Finland, however, means-tested transfers are more generous than in Sweden, and particularly so at low levels of earnings.

Unlike for Sweden, we do have disaggregated data for pensions for Finland. It is striking to note that pensions provide 36 percent of gross income for the Finnish households at the bottom 10 percent of the distribution and 22 percent for those in the bottom 25 percent. This is more than double the amounts for the U.S and Canada and about ten times the level in Germany. In Finland, 44 percent of household heads age 50 to 60 are in the bottom 10 percent of the distribution and 34 percent are in the bottom 25 percent, compared with 25 percent in the total sample. A large share of this pension income is due to public pensions. The availability and generosity of public pensions in Finland seems to encourage a large share of public employees to retire early.

The average income tax rate and its progressivity in Finland are very similar to those of Sweden. Finland implemented a tax reform in the late 1980s (Organization for Economic Cooperation and Development [OECD], 1991) that reduced marginal income tax rates while maintaining total tax revenues by broadening the income tax base and raising indirect taxes. By 1992, the highest personal income tax rate had been reduced from 51 percent to 39 percent. On the other hand, and partially offsetting this reduction, social security contributions paid by employers and employees were increased. The OECD computed that, taking increases in social security and consumption taxes into account, the effective marginal tax rate on total labor compensation did not change significantly. We do not have data on consumption or consumption taxes; therefore our computed tax payments should be considered as lower bounds of the actual ones.

**Age, early retirement, and income**

In this section we look at gross income, taxes, and transfers over the life cycle to study the relationship between age and redistribution for working age families (25 to 60 years of age).
### TABLE 7

Finnish households ranked by earnings

<table>
<thead>
<tr>
<th>Household characteristics</th>
<th>Bottom 10%</th>
<th>1st</th>
<th>2nd</th>
<th>3rd</th>
<th>4th</th>
<th>Top 10%</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Earnings, in U.S. dollars</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average earnings</td>
<td>2</td>
<td>3,722</td>
<td>22,688</td>
<td>38,214</td>
<td>69,544</td>
<td>88,581</td>
<td>33,533</td>
</tr>
<tr>
<td>Average gross income</td>
<td>17,423</td>
<td>20,473</td>
<td>31,902</td>
<td>46,374</td>
<td>76,682</td>
<td>97,418</td>
<td>43,851</td>
</tr>
<tr>
<td>Average disposable income</td>
<td>14,532</td>
<td>16,783</td>
<td>23,743</td>
<td>32,694</td>
<td>49,611</td>
<td>60,534</td>
<td>30,703</td>
</tr>
<tr>
<td>Sources of gross income, %</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Labor</td>
<td>0.0</td>
<td>15.2</td>
<td>64.5</td>
<td>75.0</td>
<td>79.3</td>
<td>76.5</td>
<td>68.0</td>
</tr>
<tr>
<td>Business</td>
<td>0.0</td>
<td>3.0</td>
<td>6.6</td>
<td>7.4</td>
<td>11.4</td>
<td>14.4</td>
<td>8.5</td>
</tr>
<tr>
<td>Cash property income</td>
<td>2.3</td>
<td>5.8</td>
<td>1.7</td>
<td>2.0</td>
<td>3.0</td>
<td>4.3</td>
<td>2.8</td>
</tr>
<tr>
<td>Total transfers</td>
<td>97.7</td>
<td>76.0</td>
<td>27.2</td>
<td>15.6</td>
<td>6.3</td>
<td>4.7</td>
<td>20.7</td>
</tr>
<tr>
<td>Social insurance</td>
<td>34.5</td>
<td>32.6</td>
<td>16.0</td>
<td>10.1</td>
<td>4.3</td>
<td>3.0</td>
<td>11.3</td>
</tr>
<tr>
<td>Means-tested</td>
<td>25.0</td>
<td>19.0</td>
<td>3.9</td>
<td>1.4</td>
<td>0.5</td>
<td>0.4</td>
<td>3.5</td>
</tr>
<tr>
<td>Pensions</td>
<td>36.5</td>
<td>22.1</td>
<td>6.0</td>
<td>3.2</td>
<td>1.0</td>
<td>0.8</td>
<td>5.0</td>
</tr>
<tr>
<td>Private</td>
<td>1.4</td>
<td>1.8</td>
<td>1.2</td>
<td>0.7</td>
<td>0.5</td>
<td>0.5</td>
<td>0.8</td>
</tr>
<tr>
<td>Income tax, %</td>
<td>13.6</td>
<td>14.7</td>
<td>19.5</td>
<td>22.6</td>
<td>27.7</td>
<td>30.3</td>
<td>23.4</td>
</tr>
<tr>
<td>Average number of earners</td>
<td>0.0</td>
<td>0.8</td>
<td>1.5</td>
<td>1.8</td>
<td>2.3</td>
<td>2.4</td>
<td>1.6</td>
</tr>
<tr>
<td>Average household size</td>
<td>1.6</td>
<td>1.8</td>
<td>2.3</td>
<td>2.8</td>
<td>3.3</td>
<td>3.4</td>
<td>2.6</td>
</tr>
<tr>
<td>Age of household head, %</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25–34</td>
<td>15.8</td>
<td>28.4</td>
<td>37.9</td>
<td>29.2</td>
<td>15.3</td>
<td>11.0</td>
<td>27.7</td>
</tr>
<tr>
<td>35–49</td>
<td>39.9</td>
<td>37.5</td>
<td>40.2</td>
<td>50.7</td>
<td>59.9</td>
<td>59.5</td>
<td>47.1</td>
</tr>
<tr>
<td>50–60</td>
<td>44.3</td>
<td>34.2</td>
<td>22.0</td>
<td>20.1</td>
<td>24.7</td>
<td>29.4</td>
<td>25.3</td>
</tr>
<tr>
<td>Average age, years</td>
<td>46.5</td>
<td>43.2</td>
<td>39.9</td>
<td>41.1</td>
<td>43.6</td>
<td>44.7</td>
<td>42.0</td>
</tr>
</tbody>
</table>


In all countries average gross income follows an inverse U-shape pattern, first increasing with age and then declining as the household head gets older (Tables 8 to 12). Total transfers follow a U-shape pattern: They are more generous for younger and older households. In fact, middle-age families on average earn more and also hold more assets. As the family gets older some of its members retire and begin receiving social security payments and pensions, therefore transfers increase. In all countries but Sweden (for which we do not have data on private pensions), total transfers to the age group 55 to 60 are actually the highest over the life cycle. The fraction of total transfers to this age group is smallest in the U.S. and Germany (11 percent), larger in Canada and Sweden (19 and 22 percent, respectively) and largest in Finland (36 percent).

The incentives to retire early in the various countries are reflected in tables 8 to 12 by the life cycle pattern of the fraction of gross income due to labor, self-employment, and total transfers. If the fraction of total transfers rises significantly for the last (or last two) age groups, while the fraction of income from labor and self-employment goes down, we have evidence that households are retiring early. The case in which transfers go up and labor income goes down while income from self-employment increases indicates that while households are reducing their labor and receiving social security and pension payments, at the same time they are engaging in some self-employment activity to supplement their income. This is more likely to happen in countries in which social security payments do not decrease sharply when people receive some extra income, at least up to some level.

The composition of total transfers and the changes in transfers as the household ages gives some indication of which programs provide more incentives toward early retirement. In a country with a social security system that has generous provisions for early retirement, we expect to see the fraction of social insurance (which includes social security payments) increase a lot for older households. In a country in which, instead, families retire early because of incentives linked to private and public pension plans, we expect the fraction of pension income to go up.

Tables 8 to 12 show that in Germany pensions are lower than in all of the other countries for all age
### TABLE 8

**Age and income in the U.S.**

<table>
<thead>
<tr>
<th>Age</th>
<th>Average gross income</th>
<th>Income sources (%)</th>
<th>Transfer sources (%)</th>
<th>Income tax</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Labor</td>
<td>Business</td>
<td>Cash property</td>
</tr>
<tr>
<td>25–29</td>
<td>28,550</td>
<td>89.4</td>
<td>3.1</td>
<td>1.6</td>
</tr>
<tr>
<td>30–34</td>
<td>37,454</td>
<td>88.7</td>
<td>4.7</td>
<td>1.4</td>
</tr>
<tr>
<td>35–39</td>
<td>44,985</td>
<td>86.2</td>
<td>6.2</td>
<td>2.7</td>
</tr>
<tr>
<td>40–44</td>
<td>48,808</td>
<td>85.6</td>
<td>6.6</td>
<td>3.1</td>
</tr>
<tr>
<td>45–49</td>
<td>54,959</td>
<td>84.2</td>
<td>7.1</td>
<td>3.8</td>
</tr>
<tr>
<td>50–54</td>
<td>54,156</td>
<td>81.2</td>
<td>7.0</td>
<td>5.1</td>
</tr>
<tr>
<td>55–60</td>
<td>49,589</td>
<td>75.4</td>
<td>6.2</td>
<td>7.2</td>
</tr>
<tr>
<td>Total</td>
<td>44,965</td>
<td>84.4</td>
<td>6.0</td>
<td>3.5</td>
</tr>
</tbody>
</table>


### TABLE 9

**Age and income in Canada**

<table>
<thead>
<tr>
<th>Age</th>
<th>Average gross income</th>
<th>Income sources (%)</th>
<th>Transfer sources (%)</th>
<th>Income tax</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Labor</td>
<td>Business</td>
<td>Cash property</td>
</tr>
<tr>
<td>25–29</td>
<td>28,707</td>
<td>84.0</td>
<td>4.1</td>
<td>0.9</td>
</tr>
<tr>
<td>30–34</td>
<td>34,120</td>
<td>82.0</td>
<td>6.4</td>
<td>1.1</td>
</tr>
<tr>
<td>35–39</td>
<td>37,227</td>
<td>82.8</td>
<td>6.1</td>
<td>1.1</td>
</tr>
<tr>
<td>40–44</td>
<td>40,909</td>
<td>82.5</td>
<td>7.6</td>
<td>1.5</td>
</tr>
<tr>
<td>45–49</td>
<td>43,849</td>
<td>82.1</td>
<td>6.6</td>
<td>2.4</td>
</tr>
<tr>
<td>50–54</td>
<td>44,607</td>
<td>81.9</td>
<td>6.0</td>
<td>2.3</td>
</tr>
<tr>
<td>55–60</td>
<td>38,593</td>
<td>69.1</td>
<td>7.3</td>
<td>4.4</td>
</tr>
<tr>
<td>Total</td>
<td>38,230</td>
<td>80.9</td>
<td>6.5</td>
<td>1.9</td>
</tr>
</tbody>
</table>


### TABLE 10

**Age and income in Germany**

<table>
<thead>
<tr>
<th>Age</th>
<th>Average gross income</th>
<th>Income sources (%)</th>
<th>Transfer sources (%)</th>
<th>Income tax</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Labor</td>
<td>Business</td>
<td>Cash property</td>
</tr>
<tr>
<td>25–29</td>
<td>31,747</td>
<td>87.7</td>
<td>1.7</td>
<td>0.3</td>
</tr>
<tr>
<td>30–34</td>
<td>39,949</td>
<td>85.8</td>
<td>7.1</td>
<td>0.9</td>
</tr>
<tr>
<td>35–39</td>
<td>45,944</td>
<td>81.3</td>
<td>10.0</td>
<td>2.3</td>
</tr>
<tr>
<td>40–44</td>
<td>48,523</td>
<td>78.7</td>
<td>12.0</td>
<td>2.7</td>
</tr>
<tr>
<td>45–49</td>
<td>53,947</td>
<td>79.4</td>
<td>11.6</td>
<td>3.1</td>
</tr>
<tr>
<td>50–54</td>
<td>62,911</td>
<td>82.7</td>
<td>5.2</td>
<td>7.0</td>
</tr>
<tr>
<td>55–60</td>
<td>45,906</td>
<td>81.4</td>
<td>3.4</td>
<td>3.7</td>
</tr>
<tr>
<td>Total</td>
<td>46,092</td>
<td>82.2</td>
<td>7.5</td>
<td>3.0</td>
</tr>
</tbody>
</table>

groups. In particular, if we compare the 55 to 60 age group, the fraction of gross income coming from pensions is 1 percent in Germany, 6 percent in the U.S., 7 percent in Canada, and a large 20 percent in Finland. Social insurance, which includes social security transfers, for the same age group represents respectively, 9 percent of gross income in Germany, 4 percent in the U.S., 7 percent in Canada, and 13 percent in Finland. Correspondingly, German families whose head is 55 to 60 are the ones with the highest fraction of gross income coming from labor: 81 percent, compared with a low of 50 percent in Finland. These numbers reflect the fact that the German system provides less incentive toward early retirement than in the other countries. At the opposite extreme is the Finnish system. In Finland, the fraction of gross income due to labor drops from 70 percent at age 50 to 50 percent for age 55 to 60. However, the fraction of income deriving from self-employment activities is higher than in the other countries and is even higher for older family heads. This indicates that in Finland people retire early and devote part of their time to self-employment.

As we discussed earlier, Sweden and Finland are the countries with the most generous transfer systems and highest average tax rates. We do not have data for private pensions in Sweden, and public pensions are included in social insurance. Looking at social insurance transfers, we see that their fraction of gross income increases from 12 percent at age 50 to 21 percent at age 55 to 60, while labor income decreases from 80 percent to 70 percent. In Sweden income from self-employment increases with age, flattening out at 2.8 percent around age 45 to 49 and
staying at that level. The available information for Sweden suggests that there are some incentives to retire early and that households do not supplement their income through self-employment to the same extent as our data suggest for Finland.

The U.S. and Canada seem to provide more incentives to retire early than Germany, but much less than Finland and Sweden. In both the U.S. and Canada, the transfer component that increases the most for the lowest or oldest income group is the pension component. The effect is somewhat stronger in Canada than in the U.S.

Conclusion

All of the various measures of income we look at are unequally distributed across countries, and their distributions are concentrated and skewed. The governments of these five countries have some commitment to reducing income inequality. However, they go about this task with different intensities and they use rather different tools to achieve it. The data for the U.S. indicate less commitment to reducing income inequality and a strong emphasis on progressive taxation as a redistribution device. Moreover, a large portion of the transfers to the poorest segment of the population are means tested.

Canada is quite close to the U.S., both in terms of size of redistribution and instruments used, with only slightly more emphasis on transfers.

Germany appears to focus on reducing labor income inequality through other policies, with less emphasis on taxes and transfers.

Sweden and Finland engage in substantial redistribution of income, using high average tax rates, little tax progressivity, and aggressive transfers. Sweden uses mainly social insurance transfers, while Finland relies a little more on means-tested transfers, but not nearly as much as the U.S. and Canada.

Our results provide some useful lessons for public policy. First, as we discussed in the introduction, economic theory suggests that there is a tradeoff between redistribution and efficiency: Transferring more income to the poorer people tends to reduce their work effort during their working years and may induce them to retire early. In addition, it can distort the economic decisions of those who are taxed to provide the revenues that are being redistributed.

Second, there are theoretical reasons why the distribution of labor income should depend on the tax and transfer system, as well as on the distribution of human capital. Human capital is linked to education, which in turn is influenced by government subsidies.

Our research provides evidence that is consistent with these theoretical propositions. It is interesting to notice that, as a result of all of these forces, the distribution of labor earnings in the countries that traditionally are more concerned with redistribution (Finland and Sweden) are not necessarily more equal than the ones that belong to the Anglo-Saxon tradition of low government intervention (the U.S. and, perhaps, Canada). Finland is one obvious example of a country of high government intervention and high labor income inequality. This is partly due to a more pronounced pattern of early retirement in Finland than in all of the other countries. Furthermore, Finland’s relatively generous unemployment benefits may discourage job search and work effort. This could translate into a larger number of unemployed or underemployed, which increases measured inequality in labor earnings.

Our findings are thus consistent with the prediction from economic theory that greater redistribution through taxes and transfers is achieved at the cost of greater distortions on labor supply and early retirement decisions.

Consistent with other theoretical work, we also find that high redistribution countries rely heavily on instruments other than income taxes, such as transfers based on special conditions or means testing, to achieve high levels of redistribution while keeping distortions as low as possible for the beneficiaries. This is costly because it generates the need to monitor eligibility.

In Germany the level of redistribution through taxes and transfers is low. However, the distribution of labor earnings is remarkably more equal than in all of the other countries we consider. Evidently, the German government is using other instruments to achieve this, possibly more equal access to public education. Another factor may be the presence of powerful labor unions, which typically support a flat wage structure that enhances security at the expense of incentives.

APPENDIX

The data

Because of either underreporting or lack of oversampling of the rich, the people at the upper tail of the earnings distribution are underrepresented in our datasets. Income from self-employment and income from interests and dividends are especially subject to underreporting.
For Germany and Finland, the original datasets did not allow the reporting of negative earnings, and set them to zero. To make our data more homogeneous across countries, we set negative earnings to zero also for the other countries.

**The U.S. dataset**

The sampling frame for the survey consists of all occupied housing units. The sampling frame is a multistage stratified probability sample of the population. Of the households participating in the survey in 1979, 8 percent to 9 percent refused to answer any of the income questions. If these cases are combined with others for which responses to some but not all income questions occurred, the “item” nonresponse rate for income amounts averages about 15 percent. Higher rates of missing responses were found for self-employment income (33 percent) and property income (25 percent). Imputation procedures were used by the CPS to replace the nonresponse to the question with an answer that was typical of other households with similar characteristics. This imputation procedure partly corrects for the bias due to the fact that nonrespondents have, on average, higher levels of income than respondents.

The CPS also compared the aggregates derived from the CPS dataset and the ones from the national income account, and adjustments were made. Even after the adjustments, property income (interest, dividends, rent) and means-tested transfer income data are of poor quality. Moreover, due to general non-sampling errors at the upper tail of the income distribution, very rich people are not well represented. The number of observations for households whose head is of working age (25–60) in the 1994 wave that we use is 41,871.

**The Canadian dataset**

The sampling frame includes all private dwellings in the ten Canadian provinces. A stratified cluster probability sample design was employed. In the 1987 Survey of Consumer Finances (SCF), 20 percent of individuals did not respond to income questions. The missing values were imputed. Some specific income items (for example, investment income sources and some government transfers) were undercovered. The top end of the income distribution curve was underrepresented in the sample. In the dataset that LIS derived from the 1994 wave of the Canadian SCF, there are 4,224 households whose head is of working age.

**The German dataset**

The sampling frame was given by the list of registered voters. The German Socio-Economic Panel employs a two-stage stratified sample design. Adjustments and corrections to the original dataset were made to improve data quality. However, the dataset still suffers from a relatively high number of missing values. To get around this problem, we dropped the households for which we did not have the information on either earnings, income or disposable income (about 8 percent of our sample). In the dataset that LIS derived from the 1994 wave for Germany there are 4,224 households whose head is of working age.

**The Swedish dataset**

The sampling frame for the Income Distribution Survey is the taxation register for all individuals 18 years of age and older. A four-stage stratified sample design was used. The sample design was used to control the sample size for farmers, employers, and pensioners. Evaluations of the quality of these income data were not performed and no corrections or adjustments were made to the original data. However, since the data come from the taxation register, there are no missing data for income. In the dataset that LIS derived from this source, there are 8,720 households whose head is of working age.

**The Finnish dataset**

The sampling frame for the Finnish Income Distribution Survey is the taxation register for the total population of household heads. As in the Swedish dataset, there are no missing data for income. Some population groups have been oversampled, such as farmers, other entrepreneurs, and other high-income groups. This is corrected through the weighting procedure. In the dataset derived by LIS from the 1995 wave there are data on 7,033 households.

---

**NOTES**

1Stokey (1999) provides an overview of the literature on intergenerational mobility in the U.S. She concludes that even in the country considered the “land of equal opportunity,” children from rich families have more chances for economic success than children from poor families.

2See Heckman, Lochner, and Taber (1998) for a theoretical model estimated on the U.S. data, in which many of these elements interact dynamically.

3See, for example, Cremer and Pestieau (1996).

4Salary income includes all forms of cash wage and salary income, including employer bonuses, gross of employee social insurance contributions/taxes but net of employer contributions/taxes.

5Typically, α is chosen to be between 0 and 1. When α = 0 we get back to the benchmark case we discuss throughout the article.
Total family income is the unit of analysis. Should one choose \( \alpha = 1 \), the unit of analysis would be per capita family income. To check our results against the case \( \alpha = 0 \), we choose \( \alpha = 5 \), which is a number commonly used in the literature.


At first, this might seem surprising because in most countries the distribution of wealth is very concentrated. In the U.S., the top 5 percent of people hold 50 percent of the total wealth, while the bottom 40 percent of people hold only 4 percent of total wealth (Wolff, 1987). As a result, income from capital is also highly concentrated. Moreover, one could expect a high correlation between wealth and earnings. Díaz-Giménez, Quadrini, and Rios-Rull (1997) find a small correlation (.23) between earnings and wealth, but include retirees in their sample. The correlation between earnings and wealth should be higher in our subsample. However, LIS does not provide data on assets so we cannot compute it.

As we discussed above, cash property income is more concentrated than earnings because the distribution of wealth itself is. This implies that when we add cash property income to earnings, this increases the fraction of total factor income held by the richest people. This would increase the Gini coefficient. However, adding cash property income also reduces the fraction of people at zero or negative wealth, thereby reducing the Gini index. In our dataset, the two forces counterbalance each other in each country so that the Gini coefficients for earnings and factor income in every country are basically the same. The fact that the Gini coefficient is unchanged is likely to be a consequence of the underreporting of interest and dividend income and of the underrepresentation of the very rich people.

We could choose different cutoffs for the comparisons. For example, the richest 10 percent with the poorest 10 percent. We choose to look at the poorest 20 percent because this is the smallest fraction of people that have positive earnings in all of the countries we consider.

See Crawford and Lilien (1981) for a theoretical paper on how social security influences retirement decisions.

We do not report the Lorenz curves for factor income across the various countries because they overlap almost perfectly with the ones for earnings and the patterns are similar to those described above. This is probably a consequence of the fact that we do not have good data on interests and dividends.

Each quartile includes 25 percent of the households in our sample, including all working age families, ordered from poorest to richest.

We do not report the information on factor income separately because in this sample its distribution is very close to the one for earnings, as we noted previously.

Examples of near-cash benefits are food stamps and housing benefits.

In the dataset for Sweden, public pensions are lumped together with social security transfers and we have no data for private pensions. As a result, our computation underestimates total transfers in Sweden.

REFERENCES


The expectations trap hypothesis

Lawrence J. Christiano and Christopher Gust

Introduction and summary

Many countries, including the U.S., experienced a costly, high inflation in the 1970s. This article reviews some research devoted to understanding why it happened and what can be done to prevent it from happening again.

We take it for granted that the high inflation was the result of high money growth produced by the U.S. Federal Reserve. But, to make sure that it does not happen again, it is not enough to know who did it. It is also necessary to know why the Fed did it. We hypothesize that the Fed was in effect pushed into producing the high inflation by a rise in inflationary expectations of the public. In the language of Chari, Christiano, and Eichenbaum (1998), we say that when a central bank is pressured to produce inflation because of a rise in inflation expectations, the economy has fallen into an expectations trap. We call this hypothesis about inflation the expectations trap hypothesis.

We argue that the dynamics of inflation in the early 1970s are consistent with the expectations trap hypothesis. We describe two versions of this hypothesis. We also describe an alternative hypothesis, which we call the Phillips curve hypothesis. According to this hypothesis, inflation occurs when a central bank decides to increase money growth to stimulate the economy and is willing to accept the risk of high inflation that that entails. The expectations trap hypothesis and the Phillips curve hypothesis both maintain that high inflation is a consequence of high money growth. Where they differ is in the motives that they ascribe to the central bank.

Much of our analysis assessing the various hypotheses about inflation is based on an informal review of the historical record. We supplement this discussion by studying a version of the expectations trap hypothesis using a general equilibrium, dynamic macroeconomic model. There are two reasons that we do this. First, we want to demonstrate that the expectations trap hypothesis can be integrated into a coherent view of the overall macroeconomy. Second, we want to document that this hypothesis has the potential to provide a quantitatively realistic account for the 1970s take-off in inflation.

The model we use is the limited participation model studied in Christiano and Gust (1999). It requires a specification of monetary policy in the 1970s, and for this we use the policy rule estimated by Clarida, Gali, and Gertler (1998). The account of the early 1970s that we produce using the model posits that a bad supply shock (designed to capture the various commodity shortages of the early 1970s) triggered a jump in expected inflation, which then became transformed into higher actual inflation because of the nature of monetary policy. We find that, consistent with the data, the model predicts stagflation. We view this result as supportive of the expectations trap hypothesis.

We compare our model with an alternative quantitative model of the 1970s inflation proposed by Clarida et al. That model can also explain the rise in inflation in the 1970s as reflecting a self-fulfilling increase in inflation expectations. It is a sticky price, rational expectations version of the IS–LM model.
What is an expectations trap?

We begin with an abstract definition of an expectations trap. We then describe two particular types of expectations traps. Finally, we ask, What is the ultimate cause of inflation under the expectations trap hypothesis?

The trap, defined

An expectations trap is a situation in which an increase in private agents’ expectations of inflation pressures the central bank into increasing actual inflation. There are different mechanisms by which this can happen. However, the basic idea is always the same. The scenario is initiated by a rise in the public’s inflation expectations. Exactly why their inflation expectations rise doesn’t really matter. What does matter is what happens next. On the basis of this rise in expectations, private agents take certain actions which then place the Fed in a dilemma: either respond with an accommodating monetary policy which then produces a rise in actual inflation or refuse to accommodate and risk a recession. A central bank that is responsive to concerns about the health of the economy could very well wind up choosing the path of accommodation, that is, falling into an expectations trap.

A cost-push trap and a working capital trap

We describe two versions of the expectations trap hypothesis, which differ according to the precise mechanism by which higher inflation expectations pressure the Fed into supplying more inflation. One mechanism, presented in Chari, Christiano, and Eichenbaum (1998), is similar to the conventional cost-push theory of inflation. We call it a cost-push expectations trap. Here is how it works. Higher inflation expectations lead people to demand, and receive, higher wage settlements. Firms are happy to pay the increased wages because, expecting a rise in the general price level, they think they can pass along the higher wage costs in the form of higher prices. This puts the Fed in the dilemma mentioned above. The Fed can produce the inflation everyone expects by raising money growth. Or, if it does not, it will put the economy through a recession. Under some circumstances, the Fed will not be willing to tolerate the recession and will feel compelled to produce inflation. In this case, the Fed ends up validating the original rise in inflation expectations. We call this hypothesis about inflation, the cost-push version of the expectations trap hypothesis.

We shall see that this version of the expectations trap hypothesis encounters some difficulties explaining the high inflation of the 1970s. We now describe another version of this hypothesis, which does not have these problems.

The limited participation model of money, which is analyzed below, highlights a different mechanism by which an expectations trap can occur. We call this a working capital expectations trap. It relies on the assumption that firms must borrow funds in advance (acquire working capital) in order to finance some or all of the inputs needed to carry on production. Under these circumstances a high nominal interest rate has a negative impact on economic activity because it raises the cost of working capital. To see how this mechanism works, suppose, again, that there is a jump in inflation expectations. Private agents, correctly perceiving that the central bank is afraid of the negative output effects of high interest rates, anticipate that the higher future inflation will be associated with low real interest rates. This leads them to cut back on saving, putting upward pressure on interest rates in the market for loanable funds. This places the central bank in a dilemma. If it keeps the money supply unchanged, then the higher expected inflation will not occur. However, the reduced saving would result in high interest rates. By drying up the supply of working capital, this would significantly slow the economy. A central bank that is concerned about the health of the private economy may prefer a second option: prevent a substantial rise in interest rates by injecting money into the economy. This has the effect of validating the initial jump in inflation expectations. Choosing this second option is another way to fall into an expectations trap. We call this hypothesis about inflation the working capital version of the expectations trap hypothesis.
Ultimate cause of inflation

Where, under the expectations trap hypothesis, does the ultimate responsibility for inflation lie? To answer this requires identifying the cause of the rise in inflation expectations. According to the expectations trap hypothesis, the cause lies with monetary institutions themselves. If, for example, the nature of those institutions is such that people cannot imagine a set of circumstances in which the central bank would accommodate a rise in inflation, then there is little reason for inflation expectations to suddenly jump. Expectations traps just couldn’t happen.

To see this, imagine there is an oil shortage. Certainly, one might reasonably expect this to lead to a rise in the price level. Because of various lags, this rise might actually take place over a period of time, maybe even a year or two. But, there is nothing in conventional economic reasoning that would connect an oil shortage to the sustained, decade-long rise in prices that we call inflation. Anyone who inferred from a 10 percent jump in the price level in one year that prices would continue jumping like this and be 100 percent higher in ten years, would be viewed as a crank. Such a person would seem as foolish as the person who, seeing the temperature outside drop one degree from one day to the next, forecasts a drop in the temperature by 100 degrees over the next 100 days.

Now consider an economy whose monetary institutions are known to assign a high priority to output and employment. In addition, suppose that that economy’s central bank has no way of credibly committing itself in advance to keeping money growth low. In a society like this, the idea that inflation could take off seems quite plausible. In such a society, even seemingly irrelevant events could spark a rise in inflation expectations. For example, a person who revised upward their inflation forecast in the wake of an oil shock would now not necessarily seem like a crank. There are a number of ways they could back up their forecast with sensible economic reasoning. Such a person could use either of the two expectations trap arguments described above.

So, the expectations trap hypothesis lays responsibility for inflation with monetary institutions. To reduce the possibility of expectations traps, the institutions must be designed so that the central bank’s commitment to fighting inflation is not in doubt. Under these circumstances, people participating in wage negotiations who profess to believe inflation is about to take off will be met with disbelief rather than a higher wage settlement.

How exactly monetary institutions should be designed to reduce the likelihood of an expectations trap is controversial. But, there is one point on which there appears to be agreement. The central banker at the very least should make a show of not being too concerned about the health of the economy. An example of this can be found in the reaction to a famous (or infamous) speech by the then vice-chairman of the Federal Reserve, Alan Blinder, at a conference in Jackson Hole, Wyoming, in 1994. In that speech, Blinder acknowledged that it is feasible for a central bank to influence unemployment and output. This generated an uproar. Many who objected probably did not do so because they thought what Blinder said was wrong. Instead, they simply thought it unwise that a central banker should let on that he thinks about such things. Why shouldn’t he let on? One possibility—the one emphasized in the expectations trap hypothesis—is that the greater the apparent concern by the central bank for the real economy, the greater is the risk of falling into an expectations trap.

Background events

We provide a brief review of the basic economic events leading up to the high inflation of the 1970s. We argue that the data appear consistent with the hypothesis that the U.S. became ensnared in an expectations trap by the late 1960s and early 1970s. We then compare the expectations trap hypothesis about inflation with another hypothesis. According to that hypothesis, the Fed consciously produced the high inflation as a necessary, though unfortunate, byproduct of its aggressive attempts to stimulate the economy. We call this the Phillips curve hypothesis, because it involves the Fed’s attempts to exploit the Phillips curve. Finally, we look at the data to identify the economic consequences of the take-off in inflation in the early 1970s.

Events leading up to the 1970s: Setting the trap

An important part of the story of the inflation of the 1970s begins with the recession of the early 1960s. That recession helped bring the administration of John F. Kennedy into power. Kennedy brought with him the best and the brightest Keynesian minds of the time. The chairman of the Council of Economic Advisers (CEA) was the very distinguished Keynesian economist, Walter Heller. Members of the CEA included another distinguished Keynesian economist, the future Nobel laureate, James Tobin. Government policy was animated by the Keynesian conviction that if the economy was performing below its potential,
then it was the responsibility of the government to use the fiscal and monetary policies at its command to restore it to strength. Figure 1 displays the federal funds rate and the growth rate of the monetary base, using annual data. Also exhibited are the years designated by the National Bureau of Economic Research to be periods of business cycle contraction (shaded area) and expansion (non-shaded area). The figure shows that the growth rate in the monetary base began to pick up in the early 1960s. The CEA also set to work to craft an expansionary fiscal policy, and one of the products of those efforts was the tax reduction legislation of 1964. Confidence in the feasibility and desirability of Keynesian stabilization policy soared with the long expansion of the 1960s.

Figure 2 shows that inflation started to pick up with a few years’ delay, in 1965. As these observations suggest, that initial rise in inflation is probably not an example of an expectations trap. It is probably best understood in terms of the Phillips curve hypothesis: It was the consequence of expansionary monetary policy, deliberately undertaken to stimulate a weak economy. It is the dynamics of inflation after the initial uptick in the 1960s that appears to take on the character of an expectations trap.

Figures 1 and 2 show that inflation proceeded to hit three peaks, one in the early 1970s, one in early 1975, and the final one in late 1980. The initial pickup in inflation in the 1960s was noted with alarm by policymakers, who responded with a very sharp rise in the federal funds rate in 1969. This policy tightening is often credited with producing the 1970 recession. Policymakers expressed dismay that the inflation rate continued to be high, even as the economy began to slide into recession (see figure 1). Arthur Burns, the chairman of the Federal Reserve at this time, said in a speech at Pepperdine College, Los Angeles, in December 7, 1970:

The rules of economics are not working in quite the way they used to. Despite extensive unemployment in our country, wage rate increases have not moderated. Despite much idle industrial capacity, commodity prices continue to rise rapidly. (Burns, 1978, p. 118)

The policy establishment became convinced that the underlying driving force of inflation was inflation expectations and that these expectations were all but impervious to recession. In a statement before the Joint Economic Committee of the U.S. Congress in 1971, Burns explained the role of inflation expectations as follows:

Consumer prices have been rising steadily since 1965—much of the time at an accelerating rate. Continued substantial increases are now widely anticipated over the months and years ahead... In this environment, workers naturally seek wage increases sufficiently large... to get some protection against...
future price advances. ... [T]houghtful em-
ployers ... reckon, as they now generally do,
that cost increases probably can be passed on
to buyers grown accustomed to inflation.
(Burns, 1978, p. 126)

Policymakers understood that, in principle, in-
fation could be stopped with a sufficiently restrictive
monetary policy, but they were concerned that the
short-run costs, in terms of lost output, would be
intolerable. In an appearance before the House of
Representatives, Committee on Banking and Currency,
July 30, 1974, Burns said:

One may therefore argue that relatively high
rates of monetary expansion have been a per-
missive factor in the accelerated pace of infla-
tion. I have no quarrel with this view. But an
effort to use harsh policies of monetary restraint
to offset the exceptionally powerful inflationary
forces of recent years would have caused seri-
ous financial disorder and economic disloca-
tion. That would not have been a sensible
course for monetary policy. (Burns, 1978)

In remarks before the Seventeenth Annual Mon-
etary Conference of the American Bankers Association,
Hot Springs, Virginia, May 18, 1970, Burns elaborat-
ed on his views about the costs of relying on money
growth alone (without, say, wage and price controls)
to reduce inflation. He thought the costs were so
large that the strategy was fundamentally infeasible
on political grounds. In his words,

There are several reasons why excessive reli-
ance on monetary restraint is unsound. First,
severely restrictive monetary policies distort
the structure of production. General mone-
tary controls, despite their seeming impartial-
ity, have highly uneven effects on different
sectors of the economy. On the one hand,
monetary restraint has relatively slight impact
on consumer spending or on the investments
of large businesses. On the other hand, the
homebuilding industry, state and local con-
struction, real estate firms, and other small
businesses are likely to be seriously handi-
capped in their operations. When restrictive
monetary policies are pursued vigorously
over a prolonged period, these sectors may
be so adversely affected that the consequenc-
es become socially and economically intoler-
able, and political pressures mount to ease up
on the monetary brakes. ...

An effort to offset, through monetary
and fiscal restraints, all of the upward push
that rising costs are now exerting on prices
would be most unwise. Such an effort would
restrict aggregate demand so severely as to
increase greatly the risks of a very serious
business recession. If that happened, the out-
cries of an enraged citizenry would probably
soon force the government to move rapidly
and aggressively toward fiscal and monetary
case, and our hopes for getting the inflation-
ary problem under control would then be
shattered. (Burns, 1978)

Policymakers were so pessimistic about the
prospects of getting inflation under control by restric-
tive monetary policy, that in August 1971 they turned
to wage and price controls.

What happened after this may seem to be an
embarrassment to the expectations trap hypothesis,
particularly the cost-push version: Money growth
continued to be high. According to the cost-push
expectations trap hypothesis, high money growth is
theFed’s response to inflationary wage and price
contracts, which are themselves driven by inflation
expectations. But, inflationary wage and price contracts
became illegal during the wage and price control pe-
riod, which lasted until 1973. So, this hypothesis seems
to predict that money growth would have been low
during the wage–price controls, not high.

The key to reconciling the expectations trap with
this high money growth lies in interest rates. Policy-
makers were convinced that wage–price controls would
not be politically feasible if interest rates were allowed
to drift up. They thought that if this happened, the
controls would be viewed as a cover for redistribut-
ing income from people earning wages and salaries
to the (typically wealthy) people who earn interest.
They feared that if this happened, then political sup-
port for the controls would evaporate, and inflation
would take off again. So, policy was directed toward
keeping the nominal interest rate about where it was
before the severe monetary tightening of 1969 (see
figure 3). It is interesting that it required such strong
money growth to keep the interest rate at this level.
A possible explanation is that this reflects the type of
portfolio decisions emphasized in the working capital
expectations trap hypothesis described earlier. That
hypothesis predicts that, in the absence of high money
growth, household portfolio decisions motivated by
concerns about future inflation would drive up the
rate of interest.

These considerations suggest to us that although
the high money growth during wage–price controls
may well be an embarrassment to the expectations
trap hypothesis, it isn’t necessarily so.

Policymakers started dismantling wage–price
controls in 1973. They were once again surprised by
the strength with which inflation took off. They had
anticipated some inflationary pressure, and they
raised rates sharply in this period (see figure 3). But,
they were surprised at just how strong the rise in
inflation was. The increase in rates was greater than
one measure of the rise in expected inflation (see figure 3). And, it just barely kept up with actual inflation (figure 4).\textsuperscript{13} Policymakers’ resolve began to fade when output and investment started to show weakness in the middle of 1973 and hours worked began to soften in late 1973. They had indicated repeatedly that they were unwilling to countenance a severe recession in the fight against inflation. Their concerns about the recessionary costs of fighting inflation seemed credible since they appeared to have been confirmed by the experience of the 1970 recession. Moreover, the 1960s and 1970s were times when governments were expected to do good things for their citizens, and hurting a subset of them for the sake of curing a social problem seemed unfair and wrong.\textsuperscript{14} In an address before the joint meeting of the American Economic Association and the American Finance Association, on December 29, 1972, Burns expressed the general sense of the time:

Let me note, however, that there is no way to turn back the clock and restore the environment of a bygone era. We can no longer cope with inflation by letting recessions run their course; or by accepting a higher average level of unemployment. ... There are those who believe that the time is at hand to ... rely entirely on monetary and fiscal restraint to restore a stable price level. This prescription has great intellectual appeal; unfortunately, it is impractical. ... If monetary and fiscal policies became sufficiently restrictive to deal with the situation by choking off growth in aggregate demand, the cost in terms of rising unemployment, lost output, and shattered confidence would be enormous. (Burns, 1978)

So, toward late 1974, policymakers reversed course and adopted a loose monetary policy, driving interest rates down sharply, to turn the economy around. Note from figures 4 and 5 that real interest rates were negative or close to zero. Of course, as the economy entered the deep 1975 recession, inflation came down substantially anyway. But, the turnaround in monetary policy then had the implication that inflation would take off again as soon as the economy entered the expansion.\textsuperscript{15} Only later, in 1978 and 1979, did the Fed turn “tough” and consciously adopt a tight monetary policy until inflation came down (see how much higher the federal funds rate went in the early 1980s, and note how it stayed up—with the exception of a brief period of weakness in mid-1980—until after the inflation rate began to fall).

We interpret these observations as being consistent with the view that by the late 1960s and early 1970s, the U.S. economy had fallen into an expectations trap. Through their words and actions, policymakers sent two clear messages to the population:

- It is technically feasible for policymakers to stop inflation.
- The costs of doing so were greater than policymakers could accept.

Under these circumstances, it was perhaps reasonable for people to expect higher inflation. When

---

**FIGURE 3**

**Federal funds rate and inflation**

![Graph showing the annual average percent of federal funds rate and inflation from 1955 to 1990.](image)

*Note: Shaded areas indicate NBER-dated recessions. Source: Based on data from Citibase.*

**FIGURE 4**

**Ex ante real rate**

![Graph showing the annual average percent of ex ante real rate from 1962 to 1998.](image)

*Note: Shaded areas indicate NBER-dated recessions. Expected inflation based on a one-month-ahead forecast of monthly CPI inflation using five-month lags in monthly inflation, four-month lags in the federal funds rate, four-month lags in the monthly growth rate in M2, and four-month lags in the premium in the return to ten-year Treasury bonds over the federal funds rate. Source: Based on data from Citibase.*
wage–price controls began to be dismantled in 1973, it would have been reasonable for the public to think that there was now nothing left standing in the way of high inflation. Inflation expectations were even stronger than before. One indication of this is that actual inflation took much longer to begin falling during the 1974 recession than it did in the 1970 recession (see figure 3). Ironically, while policymakers expressed frustration with the public for the seeming intransigence of their inflation expectations, the true cause of that intransigence may have been the nature of the monetary policy institutions themselves. This is the implication of the expectations trap hypothesis.

**Phillips curve hypothesis**

We now briefly consider the Phillips curve hypothesis about the take-off in inflation that occurred in the early 1970s. Like the expectations trap hypothesis, this hypothesis is also fundamentally monetarist in that it interprets the rise in inflation as reflecting an increase in money growth. It differs from the expectations trap hypothesis by highlighting a different set of motives on the part of the Fed. Policymakers believed the CEA estimates that output was below potential in 1971. Under the Phillips curve hypothesis, the Fed responded to this by adopting an aggressively expansionary monetary policy for the same sort of reasons that they appear to have done so in the early 1960s, to restore output and employment.

To see that the economy was below at least one measure of potential in 1991, consider the results in figures 6 and 7. Figure 6 displays quarterly data on (log) real gross domestic product (GDP) in the U.S. for the period 1966:Q1 to 1973:Q4. In addition, we report two estimates of potential GDP based on the Hodrick and Prescott (1997) filter. One is computed using data covering the period, 1948:Q1–1998:Q1. A possible problem with this is that by using currently available data we may overstate the estimate of potential GDP available to policymakers in the early 1970s. They would not have been aware of the slowdown in trend (that is, potential) GDP that started around that time (Orphanides, 1999). This motivates our second estimate of potential output, which is based only on data for the period 1948:Q1–1973:Q4. Note from figure 6 that the qualitative difference between the two estimates of potential is as expected. However, quantitatively, the difference in levels is quite small. The implied estimates of the output gap appear in figure 7. Note that the two sets of estimates virtually coincide through 1970, and then diverge a little after that. Each estimate implies that the gap in 1971 averaged around 2 percent.
The 2 percent gap was substantial by historical standards (figure 7). Still, the notion that policymakers actively solicited higher inflation as a way to fight a weak economy conflicts sharply with the words of the chief monetary policymaker, Burns. Burns was very clear about his distaste for exploiting the Phillips curve for the sake of short-term gains. He certainly accepted the notion that policy could achieve higher output by increasing inflation. After all, his fears about the consequences of fighting inflation with reduced money growth were fundamentally based on a belief in a short-term Phillips curve. His view, which corresponded to the one espoused by Milton Friedman (1968), was that attempts to exploit the Phillips curve for short-term gains would only produce more trouble in the long run. As he put it in testimony before Wright Patman’s House Committee on Banking and Currency, July 30, 1974:

> We have also come to recognize that public policies that create excess aggregate demand, and thereby drive up wage rates and prices, will not result in any lasting reduction in unemployment. On the contrary, such policies—if long continued—lead ultimately to galloping inflation, to loss of confidence in the future, and to economic stagnation. (Burns, 1978, p. 170)

It is hard to doubt the sincerity of these words. To Burns, an important lesson of the inflation of the 1970s was that price increases produced by temporary forces could lead to an intractable inflation problem later on. It would have taken an extraordinary amount of duplicity to, on the one hand, complain about the serious economic damage caused by past policy mistakes in not counteracting temporary forces, and on the other hand contribute to them himself.

**Springing the trap**

To evaluate our models, we require a simple characterization of what happened when the economy fell into the expectations trap in the early 1970s. For this, consider figures 8–10, which display the logarithm of real GDP, total hours worked in nonagricultural business, and business fixed investment, respectively. In addition, we display linear trends, computed using the data from the beginning of the sample to 1970 Q1, and extrapolated through the end of the sample. These lines draw attention to the trend change that occurred in these variables in the early 1970s. In addition, in each case we also fit a quadratic trend to the entire sample of data.

Consider the GDP data in figure 8 first. In this case, we have also included a linear trend fit to the data for the 1970s and extrapolated to the end of the sample. What is clear, by comparing the raw data with the two linear trends, is that the growth slowdown that started in the early 1970s became even more severe in the 1980s and the early 1990s. We infer from the fact that the slowdown persisted—even accelerated—in this period, that the inflation and other transient shocks that occurred in the early 1970s must have had little to do with it. Now consider hours worked in figure 9. Note how they take off beginning in the early 1970s, and how the growth rate seems to just increase continuously throughout the following decades. Again, we infer from the fact that the growth rate continued to rise after the inflation stopped that the inflation and other temporary factors in the early 1970s were not a factor in this development. Finally, note that investment shows very little trend change in the 1970s (see figure 10). After a pause during the 1974–75 recession, investment returns to its former growth path. Investment does display weakness in the late 1980s and the 1990 recession. But after that, it grows again, returning to the pre-1970s trend line by 1997.

These trend changes in hours worked and output complicate our attempts to assess alternative explanations of the inflation of the 1970s. Ideally, we would like to remove the effect on the data reflecting the factors underlying the persistent change in trend, and study the remainder. We have not found a clean way to do this. The approach we take removes a quadratic trend from each variable and assumes that the result reflects the effects of the inflation and bad supply shocks of the early 1970s. The results are displayed in figures 11–13. In the 1974–75 recession hours worked fell to around 6 percent below trend, investment was down 11 percent, and output was down 3 percent. At the same time, inflation rose from 4 percent in 1972 to 10 percent by the end of the recession. The federal funds rate went from around 4 percent in 1972 to 10 percent by the end of the recession. The episode is a classic stagflation, with inflation going up and the economy, down.

**Models**

We now report on a quantitative evaluation of the expectations trap hypothesis. For this, we need a mathematical representation of the way the central bank conducts monetary policy and of the way the private economy is put together. We describe two models of the private economy: the limited participation model of Christiano and Gust (1999) and the sticky price, IS–LM model of Clarida et al.
Monetary policy rules

There is widespread agreement that the right way to model the Fed’s monetary policy is along the lines proposed by Taylor (1993, 1999a). He posits that the Fed pursues an interest rate target, which varies with the state of the economy. A version of this policy rule was estimated using data from the 1970s by Clarida et al. They estimated that the Fed’s monetary policy causes the actual federal funds rate, \( R_t \), to evolve as follows:

1) \( R_t = \rho R_t^* + (1 - \rho)R_t^* \).

In words, \( R_t \) is a weighted average of the current target value, \( R_t^* \), and of its value in the previous period. By setting \( \rho = 0 \), the Fed would achieve its target, \( R_t = R_t^* \) in each period. It might instead prefer \( 0 < \rho < 1 \) if \( R_t^* \) exhibits more volatility than it wishes to see in the actual funds rate. The target interest rate is determined according to the following expression:

2) \( R_t^* = \text{constant} + \alpha E_t \log (\pi_{t+1}) + \gamma y_t, \quad \pi_{t+1} = \frac{P_{t+1}}{P_t} \)

where \( P_t \) is the price level, \( E_t \) is the date \( t \) conditional expectation, and \( y_t \) is the percent deviation between actual output and trend output. The estimated values of \( \rho, \alpha, \) and \( \gamma \) are 0.75, 0.8, and 0.44, respectively. We use these parameter values in our analysis.\(^{22}\)

The idea is that a tough central banker who is committed to low inflation would adopt a rule with a large value of \( \alpha \). A central banker that is less able to commit to low inflation would have a low value of \( \alpha \). Clarida et al.’s estimate for the 1970s is relatively low. The value they estimate using data after 1979 is higher, and this is a period when monetary policy is thought to have been characterized by greater commitment to low inflation. To see how much tougher monetary policy became in 1979, consider figures 4, 5, and 14. Figures 4 and 5 show that the real rate was noticeably higher in this period. Figure 14 exhibits the difference between what the federal funds rate actually was and what it was predicted to be based on equation 1. Up until 1979, these differences were on average close to zero. After 1979, the average shifts up noticeably (see the horizontal line). This indicates that the actual funds rate in that period was higher than what a policymaker following the pre-1979 rule would have allowed.

How well does this policy rule capture our observations about monetary policy in the 1970s? In one sense, it misses. We saw that there were times when the Fed was very tough, and other times when it was accommodating. We think of this policy rule...
as capturing the Fed’s behavior on average. On average, it was accommodating.

**Two models of the private economy**

We now present a brief description of the models used in the analysis. The mathematical equations characterizing both models may be found in Christiano and Gust (1999).

Consider the limited participation model first. Recall that this model emphasizes a working capital channel in the firm sector. In order to produce output in a given period, firms must borrow funds from the financial intermediary. By increasing and decreasing its injections of liquidity, the central bank can create an abundance or scarcity of those funds. The resulting interest rate fluctuations then have a direct impact on production. A scarcity of funds in the financial intermediary drives up the interest rate and induces firms to cut back on borrowing. With fewer funds with which to hire factors of production, they cut back on production. Similarly, an abundance of funds leads to a fall in the interest rate and an expansion of output.

The mechanism whereby a rise in expected inflation may lead to a rise in actual inflation in this model was sketched earlier, but we summarize it again here for convenience. When there is an increase in expected inflation (that is, $E_t \log (\pi_{t+1})$ rises) and $\alpha < 1$, this translates into a decrease in the real interest rate, $R_t - E_t \log (\pi_{t+1})$. This leads households to reduce their deposits with the financial intermediary, and has the effect of creating a scarcity of the funds available for lending to firms. Upward pressure develops on the rate of interest. In pursuing its policy of not letting the interest rate rise too much, the monetary authority must inject some liquidity into the banking system. This injection then produces a rise in prices, thus validating the original rise in inflation expectations. Since the monetary authority does permit some rise in the nominal rate of interest (that is, $\alpha > 0$), this has the effect of depressing output, employment, consumption, and investment. Thus, the limited participation model predicts that a self-fulfilling inflation outburst is associated with stagflation.

The pure logic of the model permits an inflation outburst to be triggered for no reason at all or in response to some other shock. In our modeling exercise, we treat the jump in expectations as occurring in response to a transitory, bad supply shock. Here, we have in mind the commodity supply shocks, including the oil shock, of the early 1970s.

Now consider the Clarida et al. model. In that model, a fall in the real rate of interest stimulates the interest-sensitive components of demand. The expansion of demand raises output and employment.
through a standard sticky price mechanism. In particular, firms are modeled as setting their prices in advance and then accommodating whatever demand materializes at the posted price. As output increases, the utilization of the economy’s resources, particularly labor, increases. This produces a rise in costs and these are then gradually (as the sticky price mechanism allows) passed into higher prices by firms. In this way an increase in the expected inflation rate gives rise to an increase in actual inflation, as long as $\alpha < 1$.

A feature of Clarida et al.’s model is that it does not have investment or money. The absence of investment reflects the assumption that only labor is used to produce output. Money could presumably be incorporated by adding a money demand equation and then backing out the money stock using output and the interest rate. Clarida et al. do not do this and neither do we.

Evidently, the Clarida et al. model implies that a self-fulfilling outburst of inflation is associated with a rise in employment and output. If there were no other shocks in the model, then it is clear that the Clarida et al. model would have a problem, since it would be inconsistent with the phenomenon of stagflation observed in the 1970s. However, we treat the Clarida et al. model in the same way as the limited participation model. In particular, we model the jump in inflation expectations as occurring in response to a bad supply shock. So, in principle, it might be compatible with the low output observed in the 1970s because of the bad supply shock.

### Interpreting the Taylor rule in the two models

The various hypotheses about inflation that we discuss in this article focus on the motives of policymakers. The Taylor rule summarizes their decisions, and is silent on what motives produced these decisions. Still, in assessing the limited participation and Clarida et al. models, it is useful to speculate on what sort of motives might produce a Taylor rule with $\alpha < 1$ in these models.

In the limited participation model, we interpret $\alpha < 1$ as reflecting the working capital expectations trap considerations discussed above. That is, in this model a rise in inflation expectations confronts the Fed with a dilemma because it places the goals of low inflation and stable output in direct conflict. An interpretation of $\alpha < 1$ is that this reflects the Fed’s relatively greater concern for the output goal, as in the working capital expectations trap scenario.

By contrast, in the Clarida et al. model a rise in expected inflation does not put the low inflation, stable output goals in conflict. By simply saying no to high money growth and inflation, the Fed in the Clarida et al. model prevents output and inflation from simultaneously going above trend. So, $\alpha < 1$ in the Clarida et al. model does not appear to reflect the type of central bank dilemmas that are at the heart of the expectations trap scenarios described above. Perhaps the only interpretation of $\alpha < 1$ in the Clarida et al. model is that it reflects a mistake on the part of policymakers. Under this interpretation, policymakers were not aware that with $\alpha < 1$, a self-fulfilling inflation outburst is possible. That is, policymakers simply did not know that they could have gotten out of the high inflation by raising the rate of interest sharply. Our reading of the policymaking record of this period makes us deeply skeptical of this idea.

### Evaluating the models

Neither of our models captures the events at the level of detail described earlier, nor would we want them to. The question is whether we have a model that captures the broad outlines of the take-off in inflation in the 1970s.

We construct a simulation of the 1970s using the two models described in the previous section. We specify that the fundamental exogenous shock in this period is a shift down in the production function by 1 percent. That is, for each level of the inputs, output falls by 1 percent. Inflation expectations in the wake of this shock are not pinned down. They are exogenous variables, like the technology shock. We picked the expectations subject to two constraints. First, we
required that the limited participation model display a long-lasting, substantial response of inflation to the shock. Second, we required that the price in the period of the production function shock be the same between the two models.

Consider the limited participation model first. Figure 15 exhibits the response of the variables in that model to a bad technology shock. The shock occurs in period 2. Not surprisingly, in view of our earlier discussion, the shock drives output and employment down and inflation up. The monetary authority reacts immediately to the increase in inflation expectations by reducing the money supply to push up the rate of interest (recall, the coefficient on expected inflation in the Taylor rule is positive).

Notice the variable, $Q$, in the model. That is the part of households’ financial wealth that they hold in the form of transactions balances. When inflation expectations go up and $\alpha < 1$, then households increase $Q$ and correspondingly reduce the part of their financial wealth that they deposit with financial intermediaries. The increased value of $Q$ in period 3 reflects households’ higher inflation expectations. They understand that the monetary authority’s policy rule implies that the nominal rate of interest will go up, but that it will go up by less than the increase in inflation expectations (that is, $0 < \alpha < 1$). That is, they expect the real rate to go down. This leads them to increase the funds allocated to the goods market by raising $Q$, that is, to drain funds from the financial intermediary. To guarantee that the rate of interest only rises by a small amount ($\alpha$ is small), the monetary authority must inject funds into the financial intermediary to make up for the loss of funds due to the rise in $Q$.

The rise in the interest rate that occurs with all this produces a fall in output and employment. The stagnation persists for a long time. Money growth, inflation, and the nominal interest rate remain high for years. Output, employment, consumption, and investment are down for years. Investment is low, despite the low real rate of interest, because inflation acts like a tax on investment in this model. Note that the effects are quite large. Output and employment remain 2 percent below trend for a long time, and money growth, inflation, and interest rates are more than 6 percentage points above their steady state. The fall in investment is over 6 percent. Inflation rises from 4 percent to about 10 percent and the interest rate rises from about 7.2 percent to 10 percent. These results are tentative, however, since the size of the supply shock, 1 percent, was not based on a careful analysis of the data. Nor was the response of inflation expectations chosen carefully. Still, the results build confidence that the working capital expectations trap hypothesis can deliver quantitatively large effects.

What is the reason for these persistent and large effects following a technology shock? Fundamentally, it is bad monetary policy. With a less accommodating monetary policy, it would not be an equilibrium for inflation expectations to jump so much, and so the nominal interest rate would not rise so much. With a smaller interest rate rise, the negative output and employment response to a bad technology shock would be reduced. Figure 16 exhibits what happens in our benchmark limited participation model when the policy rule estimated by Clarida et al. to have been followed in the post-Volcker period is used. In this case, the equilibrium is (locally) unique. Note that the fall in output and employment is smaller here.

The rise in the interest rate is smaller too.

We think of a small value of $\alpha$ in the pre-Volcker policy rule as reflecting that the rule is the decision of a policymaker without an ability to commit to low inflation. If we interpret the inability to commit as reflecting that the policymaker has too soft a heart for economic agents, then there is plenty of irony here. The soft-hearted policymaker in the end does greater damage to the economy than a hard-hearted one who can commit to low inflation.

Now consider the Clarida et al. model. Figure 15 exhibits the dynamic response of the variables in that model to a 1 percent drop in technology. Note from the figure that in the Clarida et al. model, employment and output rise in response to the shock. After four quarters, output is down, but the employment response remains up for several years. This dynamic response pattern reflects two things. First, in sticky price models the direct effect on output of a bad technology shock is at most very small, since output is demand determined. As a result, a bad technology shock actually has a positive effect on employment in these models (see Gali, 1999, and Basu, Fernald, and Kimball, 1999). Second, a self-fulfilling rise in inflation by itself produces a rise in output and employment in the Clarida et al. model, as the fall in the real rate of interest stimulates the interest sensitive components of aggregate demand.

The simulation results in effect present the combined effects of both a self-fulfilling rise in inflation and a bad technology shock. In view of the observations in the previous paragraph, it is not surprising that the response of employment is positive. Output is also high for several quarters, although it eventually goes negative as the effect of the bad technology shock swamps the effect of the increase in employment. The employment response in particular puts
FIGURE 15
Response to technology shock in two different models

A. Money growth
annualized percentage rate

B. Employment
percent deviation from steady state

C. Real interest
annualized percentage rate

D. Consumption
percent deviation from steady state

E. Inflation
annualized percentage rate

F. Output
percent deviation from steady state

G. Nominal interest
annualized percentage rate

H. Investment
percent deviation from steady state

I. Transactions balances ($/c81/c116$)
percent deviation from steady state

Note: Shock happens in quarter two.
Response to a negative technology shock under two different Taylor rules

A. Money growth
annualized percentage rate

B. Employment
percent deviation from steady state

C. Real interest
annualized percentage rate

D. Consumption
percent deviation from steady state

E. Inflation
annualized percentage rate

F. Output
percent deviation from steady state

G. Interest
annualized percentage rate

H. Investment
percent deviation from steady state

I. Transactions balances (Q)
percent deviation from steady state

Note: Shock happens in quarter two.
this model in sharp conflict with the observed stagflation of the 1970s.

We conclude that the limited participation model provides a reasonable interpretation of the take-off in inflation in the 1970s as a working capital expectations trap. The effects in the model are large, and qualitatively of the right type: The model predicts a stagflation. The alternative model that we examine, the one proposed in Clarida et al., provides a less convincing explanation of the 1970s. The model predicts a boom. In addition, as discussed in the previous section, the model’s explanation of why policymakers allowed the inflation rate to take off is not very compelling.

**Conclusion**

We have argued that the expectations trap hypothesis helps explain the high inflation in the early 1970s, particularly the take-off that began in 1973. We have argued against another hypothesis, the Phillips curve hypothesis. According to that, the high inflation was an unfortunate but necessary risk that the Fed was willing to take when it decided to jump start a weakened economy in the early 1970s. These hypotheses are in fact quite similar, and so it may appear that we are splitting hairs in trying to differentiate between them. Is there anything at stake in the distinction?

We believe there is. Under the Phillips curve hypothesis, preventing a repeat of the high inflation of the 1970s is a relatively easy task: just say no to high money growth as a way to stimulate the economy. Under the expectations trap hypothesis, the problem of inflation is not solved so easily.

According to the expectations trap hypothesis, high inflation is the Fed’s reaction to pressures originating in the private economy. The entire policymaking establishment, when confronted with these pressures, may truly not want to say no. To see this, imagine that bad supply shocks drove prices and unemployment up, and people responded by signing inflationary wage and price contracts. Certainly, the Fed would not be happy about following the path of accommodation and validating the expectations incorporated in the wage and price contracts. But, it may well choose to do so anyway. With the White House, the Congress, and the public at large bearing down on it like a great tsunami, the Fed may simply feel it has no choice.

So, the expectations trap hypothesis implies that it is not so easy to prevent a resurgence of a 1970s style inflation. According to that hypothesis, fundamental institutional change is needed to guarantee that people would never reasonably expect a take-off in inflation in the first place. What sort of institutional change might that be?

We have not attempted to answer this question. There is a large range of possibilities. One is that the necessary changes have already occurred. According to that, the simple memory of what happened in the inflation of the 1970s is enough to stay the hand of a policymaker tempted to validate the expectations incorporated in inflationary wage and price contracts. This is of course an attractive possibility, but there is reason to doubt it. When the expectations trap argument is worked out formally, it is assumed that the policymaker has unlimited memory, a clear understanding of the consequences of alternative actions, and excellent foresight (see Chari, Christiano, and Eichenbaum, 1998). The logic of expectations traps simply has nothing to do with ignorance. So, the notion that expectations traps became less likely when our eyes were opened by the experience of the 1970s does not seem compelling.

Another possibility is that changes in legislation are needed, changes that focus the legal mandate of the Fed exclusively on inflation. This would make it harder for a Congress and White House, panicked by high unemployment and inflation, to pressure the Fed into tossing inflation objectives to the wind in favor of unemployment. Understanding this in advance, the public would be unlikely to raise inflation expectations in response to transient events, as it seems to have done in the early 1970s.

The expectations trap hypothesis does not say what change is needed to prevent a self-fulfilling take-off in inflation expectations. What it does say is that if the government finds a way to credibly commit to not validating high inflation expectations, then costly jumps in inflation expectations will not occur in the first place.

**APPENDIX**

**Burns and Nixon**

It has been argued that, as chairman of the Federal Reserve, Arthur Burns simply did what President Nixon told him to do. Burns initially joined the Nixon administration as a special advisor to President Nixon when the latter took office in 1968. The idea is that the boss–employee nature of that relationship continued when Nixon appointed Burns to be chairman of the Federal Reserve. This impression was reinforced by Stanford Rose in a famous article in *Fortune* magazine in 1974, which suggested that Nixon was able to interrupt the policymaking committee of the Fed.
with a one-hour telephone call and control the outcome of the meeting.

Nixon apparently did have hopes of influencing Burns when he appointed Burns chairman of the Federal Reserve. In his fascinating biography of Burns, Wells (1994, p. 42) quotes Nixon as having said to Burns: “You see to it: No recession.”

But, according to Wells (1994), the impression that Burns operated at the behest of Nixon is in fact completely untrue. Burns was a man with legendary self-confidence and a powerful, imposing personality. He had been an influential chairman of the Council of Economic Advisers under Eisenhower and left a stamp on that institution that is felt even today. During that time, according to Wells (p. 29), Burns’ relationship to Nixon was that of a “… senior partner: He was older than Nixon and enjoyed more influence with Eisenhower and his lieutenants than did the vice president. Burns thought of Nixon as a protege and treated him with what one friend described as ‘slight condescension.’ …” After Nixon became president, Burns had trouble adjusting to a subordinate position. … He lectured Nixon on whatever issue was at hand, usually at great length and in considerable detail.

Burns would also bluntly contradict the president or anyone else in the administration with whom he disagreed …” The diaries of H. R. Haldeman (1994), Nixon’s chief of staff, confirm this impression of a self-assured Burns who expected to get his way. For example, here are a couple of entries about Burns while he was in the Nixon White House: (p. 54) “… Huge Burns flap because he didn’t get in to see [the President]…,” (p. 59) “Big flap with Arthur Burns on AID …”

Wage and price controls were a major source of friction between Burns and Nixon: Burns concluded that they were necessary, and Nixon was opposed. For example, according to Haldeman (1994, p. 310) Nixon told his cabinet on June 29, 1971, “Our decisions are that there will be no wage–price controls, no wage–price board.” According to Wells (pp. 70–77), the disagreement provoked ‘ugly’ confrontations between Burns and the White House, as Burns went public with his views. In the end, in mid-August, Nixon decided to impose wage–price controls after all. The episode shows that, as Wells (1994) puts it (p. 100), “The chairman was clearly no plant tool of the chief executive but rather did whatever he thought was best.”

NOTES

1Also, see Chari, Christiano, and Eichenbaum (1998).
2This model is a modified version of the model in Christiano, Eichenbaum, and Evans (1998).
3The model is derived from a dynamic general equilibrium model with maximizing agents and cleared markets. The possibility that such a model could, under the sort of policy estimated by Clarida et al. using data from the 1970s, have an equilibrium in which inflation expectations can be self-fulfilling was first discovered by Kerr and King (1996).
4In this article, we focus on expectations traps in which inflation is high. The opposite—an expectation trap in which inflation is low—is also a possibility.
5The cost-push expectations trap is very close to the hypothesis Blinder advances as an explanation of the takeoff of inflation in the early 1970s:

Inflation from special factors can “get into” the baseline rate if it causes an acceleration of wage growth. At this point policymakers face an agonizing choice—the so-called accommodation issue. To the extent that aggregate nominal demand is not expanded to accommodate the higher wages and prices, unemployment and slack capacity will result. There will be a recession. On the other hand, to the extent that aggregate demand is expanded (say, by raising the growth rate of money above previous targets), inflation from the special factor will get built into the baseline rate. (Blinder, 1982, p. 264)
6For one prominent commentator who takes this position, see Barro (1996, pp. 58–60).
7The data are taken from Citibase. The mnemonic for the federal funds rate is fiff, and the mnemonic for the monetary base is fmbase.
8Inflation is measured as the annual percent change in the Consumer Price Index with Citibase mnemonic, pnew (CPI-W: all items).
9In the same speech, Burns showed some foresight in warning about another danger associated with the strategy of relying on reduced money growth to stop inflation. He was concerned that the nature of the lags in monetary policy were such that the variance of inflation and money growth would go up in a “stop-and-go” process.

{[The effects of monetary restraint on spending often occur with relatively long lags. ... Because the lags tend to be long, there are serious risks that a stabilization program emphasizing monetary restraint will have its major effects on spending at a point in time when excess demand has passed its peak. The consequence may then be an excessive slowdown of total spending and a need to move quickly and aggressively toward stimulative policies to prevent a recession. Such a stop-and-go process may well lead to a subsequent renewal of inflationary pressures of yet greater intensity. (Burns, 1978)}

10Money growth in 1970–74 was 5.32 percent, 7.60 percent, 7.27 percent, 8.75 percent, and 7.99 percent, respectively. The number for period $t$ is \(100 \times \text{log}(m(t)/m(t-1))\). Where $m(t)$ denotes the monetary base, $t = 1970$, 1971, 1972, 1973, and 1974.
11We address the potential for the Phillips curve hypothesis to explain high money growth during the period of wage–price controls in the next subsection.
We calculated expected inflation for figure 4 based on a one-month-ahead forecast of monthly CPI inflation using five-month lags in monthly inflation, four-month lags in the federal funds rate, four-month lags in the monthly growth rate in M2, and four-month lags in the premium in the return to ten-year Treasury bonds over the federal funds rate. The rise in real rates reported in figures 4 and 5 would have been somewhat larger if we had used the GDP deflator to measure inflation.

With the experience of the Great Depression and the intellectual foundations provided by Keynes’ General Theory, it was generally accepted that governments’ responsibility was to preserve the health of the economy. This was put into law in the Employment Act of 1946, which created the Council of Economic Advisers.

The output gap is measured as \(100 \times (\log{GDP} - \log{GDP^{1973}})\), where \(\log{GDP^{1973}}\) is the trend in log GDP implied by the HP filter.

The average gap for 1971 was \(-1.75\) percent according to the sample estimate and \(-1.99\) percent according to the sample estimate of the difference between [actual output and potential output] reports of the output gap. For another discussion of the stop-and-go success of stabilization policy in the 1960s, see Barsky and Kilian (2000). The output gap is measured as \(100 \times (\log{GDP} - \log{GDP^{1973}})\), where \(\log{GDP^{1973}}\) is the trend in log GDP implied by the HP filter. There are other output gap measures based on a different notion of trend. In these, the trend corresponds to the “nonaccelerating inflation” level of the variable: the level which, if it occurred, would produce a forecast of zero change in the rate of inflation in the near future. Gap concepts like this are fundamentally multivariate. To see how the HP filter can be adapted to correspond more closely to this alternative gap concept, see Laxton and Tello (1992) and St-Amant and Van Norden (1997). We assume that, for our purposes, it does not matter significantly whether the output gap is measured based on the adjusted or unadjusted versions of the HP filter.

The trend implicit in the HP filter is a fairly standard way to estimate potential GDP. For example, the OECD (1999, p. 205) reports estimates of the output gap computed in this way. Taylor (1999b) also uses this method to compute the output gap. Finally, according to Orphanides and van Norden (1999, p. 1), “the difference between [actual output and potential output] is commonly referred to as the business cycle or the output gap (italics added).”

For an analysis of the statistical properties of this way of computing the output gap, see Christiano and Fitzgerald (1999).

There are other output gap measures based on a different notion of trend. In these, the trend corresponds to the “nonaccelerating inflation” level of the variable: the level which, if it occurred, would produce a forecast of zero change in the rate of inflation in the near future. Gap concepts like this are fundamentally multivariate. To see how the HP filter can be adapted to correspond more closely to this alternative gap concept, see Laxton and Tello (1992) and St-Amant and Van Norden (1997). We assume that, for our purposes, it does not matter significantly whether the output gap is measured based on the adjusted or unadjusted versions of the HP filter.

The output gap is measured as \(100 \times (\log{GDP} - \log{GDP^{1973}})\), where \(\log{GDP^{1973}}\) is the trend in log GDP implied by the HP filter.

The average gap for 1971 was \(-1.75\) percent according to the full sample estimate and \(-1.99\) percent according to the sample that stops in 1973 Q4.

17To some extent, the rise in inflation was due to the oil shock in late 1973. However, about three-quarters of the price increases of that year occurred before the Yom Kippur war and the October oil embargo. The take-off in inflation in 1973 may, in part, have reflected the delayed response of prices to the high money growth that occurred during the period of wage-price controls. We attempted to estimate what fraction of the 1973 price rise reflected past money growth, but found that statistical uncertainty is too large to draw a definite conclusion.

18We calculated expected inflation for figure 4 based on a one-month-ahead forecast of monthly CPI inflation using five-month lags in monthly inflation, four-month lags in the federal funds rate, four-month lags in the monthly growth rate in M2, and four-month lags in the premium in the return to ten-year Treasury bonds over the federal funds rate. The rise in real rates reported in figures 4 and 5 would have been somewhat larger if we had used the GDP deflator to measure inflation.

19With the experience of the Great Depression and the intellectual foundations provided by Keynes’ General Theory, it was generally accepted that governments’ responsibility was to preserve the health of the economy. This was put into law in the Employment Act of 1946, which created the Council of Economic Advisers.

There is hereby created in the Executive Office of the President a Council of Economic Advisers to formulate and recommend national economic policy to promote employment, production, and purchasing power under free competitive enterprise.

See DeLong (1995) for a discussion of the post-WWII intellectual climate regarding the proper role of government in the economy and the sharp contrast with the pre-WWII climate. As noted earlier, the feasibility of the notion that the government ought to intervene in the economy seemed to be confirmed with the apparent success of stabilization policy in the 1960s.

This was precisely the stop-and-go process that Burns feared, as mentioned in note 9. For another discussion of the stop-and-go nature of inflation in this period, see Barsky and Kilian (2000).

The trend implicit in the HP filter is a fairly standard way to estimate potential GDP. For example, the OECD (1999, p. 205) reports estimates of the output gap computed in this way. Taylor (1999b) also uses this method to compute the output gap. Finally, according to Orphanides and van Norden (1999, p. 1), “the difference between [actual output and potential output] is commonly referred to as the business cycle or the output gap (italics added).” For an analysis of the statistical properties of this way of computing the output gap, see Christiano and Fitzgerald (1999).

There are other output gap measures based on a different notion of trend. In these, the trend corresponds to the “nonaccelerating inflation” level of the variable: the level which, if it occurred, would produce a forecast of zero change in the rate of inflation in the near future. Gap concepts like this are fundamentally multivariate. To see how the HP filter can be adapted to correspond more closely to this alternative gap concept, see Laxton and Tello (1992) and St-Amant and Van Norden (1997). We assume that, for our purposes, it does not matter significantly whether the output gap is measured based on the adjusted or unadjusted versions of the HP filter.

The output gap is measured as \(100 \times (\log{GDP} - \log{GDP^{1973}})\), where \(\log{GDP^{1973}}\) is the trend in log GDP implied by the HP filter.

The average gap for 1971 was \(-1.75\) percent according to the full sample estimate and \(-1.99\) percent according to the sample that stops in 1973 Q4.

20See Wells (1994), p. 72, for a further discussion of Burns’ view about the Phillips curve.

21It has been argued that even if Burns was not himself duplicitous, President Nixon was, and Burns acted at the behest of Nixon. To us, the record is inconsistent with this view. See the appendix.

22The limited participation model that we use is a modified version of the model in Christiano, Eichenbaum, and Evans (1998).

23Clarida et al. (1998) use revised data to estimate the policy rule for the 1970s. Orphanides (1997) argues that constructing \(\gamma\), using final revised data may give a very different view of \(\gamma\) than policymakers in the 1970s actually had. As noted above, he argues that the productivity slowdown that is thought to have occurred beginning in the early 1970s was not recognized by policymakers until much later in that decade. As a result, according to Orphanides, real-time policymakers in the 1970s thought that output was further below potential than current estimates suggest. In private communication, Orphanides has informed us that when he uses real-time data on \(\gamma\) and the other variables to redo the Clarida et al. estimation procedure, he finds that the point estimates for \(\alpha\), \(\alpha\), and \(\beta\) for the 1970s change. They move into the region where our models no longer imply that self-fulfilling inflation take-offs are possible. The standard errors on the point estimates are large, however, and a standard confidence interval does not exclude the Clarida et al. point estimates that we use.

24Woodford (1998) develops an alternative interpretation of \(\alpha < 1\) by building on the assumption that fiscal policy (something we abstract from in our analysis) was “non-Ricardian” during the 1970s. Using the fiscal theory of the price level, he argues that with fiscal policy satisfying this condition, the Fed was forced to set \(\alpha < 1\) to avoid an even more explosive inflation than the one that actually occurred. For a simplified explanation of this argument, see Christiano and Fitzgerald (2000). The fiscal theory of the price level offers another potential explanation of the take-off in inflation in the 1970s, one that is not based on self-fulfilling expectations and that assigns a central role to fiscal policy rather than monetary policy. While this interpretation is controversial, it deserves serious consideration. See Cochrane (1998) and Woodford (1998) for further discussion.

25The production function is \(Y_t = \exp(z_t)K_t^{1/\gamma}I_t^{1-\gamma}\), where \(Y_t\) denotes gross output, \(K_t\) denotes the stock of capital, and \(I_t\) denotes labor. The state of technology, \(z_t\), evolves according to \(z_t = \rho(z_{t-1} + \varepsilon_{t-1})\), with \(\rho = 0.95\). In the limited participation model, \(\theta = 0.36\) and in Clarida et al., \(\theta = 0\). The simulation involves setting \(\varepsilon_{t-1} = -0.01\) for \(t = 2\) and \(\varepsilon_{t-1} = 0\) for all other \(t\). With this value of \(\rho\), the state of technology remains 0.7 percent below trend after ten periods and 0.4 percent below trend after 20 periods.

26There is one important difference. Shocks to the production function can occur for any parameter values of the model. Shocks to expectations can only exist for certain parameter values.

27For details of model parameterization, see Christiano and Gust (1999). The version of the limited participation model underlying the calculations in figure 15 is the one in which investment is a cash good, what Christiano and Gust (1999) call the “benchmark” model. They also consider the version of the model in which investment is a credit good. The simulation of the 1970s using the Clarida et al. estimated Taylor rule resembles the results in figure 15.

28Feldstein (1997) has argued that high inflation hurts investment, though he emphasizes a mechanism that operates through the explicit tax system.
REFERENCES


This uses a larger value of \( \alpha \).

The result that raising \( \alpha \) above unity eliminates expectations traps (at least, locally) is somewhat model specific. In some models this does not work and the central bank would have to adopt a different policy to rule out expectations traps.

It deserves repetition that the policy rules have not been derived from well-specified optimization problems of policymakers and that our discussion represents an informal interpretation. For an explicit analysis based on policymaker optimization, see Chari, Christiano, and Eichenbaum (1998).

The reasoning is simple. Let \( D \) denote demand and \( P \) and \( Y \) denote price and output. Then, \( PY = D \). In a sticky price model, \( P \) cannot change so that if \( D \) does not change then \( Y \) cannot change either, even if there is a shock to technology. Of course, if the shock is such that it takes more people to produce a given level of output, then a fall in technology results in a rise in employment. This response of employment to a bad technology shock is not robust to all specifications of monetary policy. For example, if \( \alpha \) is sufficiently large in the Clarida et al. model, then the rise in anticipated inflation produced by a bad technology shock leads the monetary authority to raise the interest rate a lot, driving down \( D \). If the fall in \( D \) is sufficiently large, then a bad technology shock could actually lead to a fall in employment. Our results indicate that under the estimated monetary policy rule, employment rises after a bad technology shock in the Clarida et al. model.


Subordinated debt as bank capital: A proposal for regulatory reform

Douglas D. Evanoff and Larry D. Wall

Introduction and summary
Last year, a Federal Reserve Study Group, in which we participated, examined the use of subordinated debt as a tool for disciplining bank risk taking. The study was completed prior to the passage of the 1999 U.S. Financial Services Modernization Act and the results are reported in Kwast et al. (1999). The report provides a broad survey of the academic literature on subordinated debt and of prevailing practices within the current market for subordinated debt issued by banking organizations. Although the report discusses a number of the issues to be considered in developing a policy proposal, providing an explicit proposal was not the purpose of the report. Instead, it concludes with a call for additional research into a number of related topics.

In this article, we present a proposal for the use of subordinated debt in bank capital regulation. Briefly, our proposal would require that banks hold a minimum level of subordinated debt and be required to approach the marketplace on a somewhat regular basis to roll over that debt. We believe the proposal is particularly timely for a variety of reasons, one of which is that Congress recently demonstrated its interest in the topic when it passed the U.S. Financial Services Modernization Act (Gramm-Leach-Bliley Act). The act instructs the Board of Governors of the Federal Reserve and the Secretary of the Treasury to conduct a joint study of the potential use of subordinated debt to bring market forces to bear on the operations of large financial institutions and to protect the deposit insurance funds. The act also requires large U.S. national banks to have outstanding (but not necessarily subordinated) debt that is highly rated by independent agencies in order to engage in certain types of financial activities. Another reason to consider alternatives now is that banks in most developed countries, including the U.S., are relatively healthy. This reduces the probability that a greater reliance on market discipline will cause a temporary market disruption. Additionally, history shows that introducing reforms during relatively tranquil times is preferable to being forced to act during a crisis.

Perhaps the most important reason that now may be a good time to consider greater reliance on subordinated debt is that international efforts to reform existing capital standards are highlighting the weaknesses of the alternatives. In 1988, the Basel Committee on Banking Supervision published the International Convergence of Capital Measurement and Capital Standards, which established international agreement on minimum risk-based capital adequacy ratios. The paper, often referred to as the Basel Capital Accord, relied on very rough measures of a bank’s credit risk exposure, however, and banks have increasingly engaged in regulatory arbitrage to reduce the cost of complying with the requirements (Jones, 2000). The result is that by the end of the 1990s, the risk-based capital requirements had become more of a compliance issue than a safety and soundness issue for the largest and most sophisticated banks.

Bank supervisors have recognized the problems associated with the 1988 accord, and the Basel Committee recently proposed two possible alternatives: a standardized approach that uses credit rating agencies to evaluate individual loans in banks’ portfolios and an internal ratings approach that uses the ratings of individual loans that are assigned by banks’ internal ratings procedures. An important element of both...
of these proposals is that they rely on risk measures obtained from private sector participants rather than formulas devised by supervisors. The use of market risk measures has the potential to provide substantially more accurate risk measurement than would any supervisory formula. Market participants have the flexibility to evaluate all aspects of a position and assign higher risk weights where appropriate.

Whether either of these approaches would result in a significant improvement, however, is questionable. The approaches share two significant weaknesses. First, both ask for opinions rather than relying on private agents’ behavior. Economists have long been trained to focus on prices and quantities established in arms-length transactions rather than on surveys of individual opinions. The problem with opinions is that individuals’ responses may depend not only on their beliefs but also on what they want the questioner to think. Second, the reliance in this case on opinions is especially problematic because the two parties being asked about a bank’s risk exposure both have an incentive to underestimate that exposure. The firm seeking a rating compensates the ratings agencies. If the primary purpose of the rating is to satisfy bank supervisors, then firms will have a strong incentive to pressure the agencies to supply higher ratings. The incentive conflict for banks is even more direct. The intent of Basel’s capital proposal appears to be to require banks to hold more capital than they otherwise would. If this is true, banks will have incentives to systematically underestimate their risk exposure.

The use of a risk measure obtained from the subordinated debt market has the potential to avoid both of these problems. The measure could use actual prices rather than some individual’s opinion. Further, the interests of subordinated debt creditors are closely aligned with those of bank supervisors, in that subordinated creditors are at risk of loss whenever a bank fails.

Below, we summarize some of the existing subordinated debt proposals. Then, we introduce our new proposal, address some of the common concerns raised about the viability of subordinated debt proposals, and explain how our proposal addresses these concerns.

**Brief summary of past proposals**

Since the mid-1980s there have been a number of regulatory reform proposals aimed at capturing the benefits of subordinated debt (sub-debt). Below, we provide a partial review of previous proposals that emphasizes the characteristics on which our proposal rests. (These are surveyed in greater detail in Kwast et al., 1999). It was common in the earlier proposals for the authors not to provide a comprehensive plan, but instead to stress the expected benefits and describe how these could be realized. Specific characteristics were typically excluded to avoid having the viability of the proposals determined by the acceptance of the details. The typical benefits of the proposals relate to the ability of sub-debt to provide a capital cushion and to impose both direct and derived discipline to banks and from the tax benefits of debt. These benefits include the following:

- a bank riskiness or asset quality signal for regulators and market participants,
- a more prompt failure resolution process, resulting in fewer losses to the insurance fund,
- a more methodical failure resolution process because debtholders unlike demand depositors must wait until the debt matures to “walk” away from the bank rather than run, and
- a lower cost of capital because of the tax advantages of deducting interest payments on debt as an expense, enabling banks to reduce their cost of capital and/or supervisors to increase capital requirements.

Horvitz (1983, 1984) discusses each of these advantages in his initial sub-debt proposal and extends that discussion in Benston et al. (1986). He challenges the view that equity capital is necessarily preferable to debt. While equity is permanent and losses can indeed be charged against it, he questions why one would want to keep a troubled bank in operation long enough to make this feature relevant. Similarly, while interest on debt does represent a fixed charge against bank earnings, whereas dividends on equity do not, a bank with problems significant enough to prevent these interest payments has most likely already incurred deposit withdrawals and has reached, or is approaching, insolvency. Arguing that higher capital levels are needed at the bank level and are simply not feasible through equity alone, Horvitz states that sub-debt requirements of “say, 4 percent of assets” are a means to increase total capital requirements to 9 percent to 10 percent. Without providing specifics, he argues that debtholders would logically require debt covenants that would give them the right to close or take over the bank once net worth was exhausted. Thus, sub-debt is seen as an ideal cushion for the Federal Deposit Insurance Corporation (FDIC).

Keehn (1988) incorporates sub-debt as a centerpiece of the comprehensive “FRB-Chicago Proposal” for deregulation. The plan calls for a modification of the 8 percent capital requirement to require that a minimum of 4 percent of risk-weighted assets be held as sub-debt. The bonds would have maturities
of no less than five years, with the issues being staggered to ensure that between 10 percent and 20 percent of the debt would mature and be rolled over each year. A bank’s inability to do so would serve as a clear signal that it was in financial trouble, triggering regulatory restrictions and debt covenants. Debt covenants would enable the debtholders to initiate closure procedures and would convert debtholders to an equity position once equity was exhausted. They would have a limited time to recapitalize the bank, find a suitable acquirer, or liquidate the bank. Keehn argues that debtholders could be expected to effectively discipline bank behavior and provide for an orderly resolution process when failure did occur. The discipline imposed by sub-debt holders could differ significantly from that imposed by depositors as holders of outstanding sub-debt could not run from the bank, but could only walk as issues matured. The potential for regulatory forbearance is also thought to be less as holders of sub-debt would be less concerned with giving the troubled bank additional time to “correct” its problems and would pressure regulators to act promptly when banks in which they had invested encountered difficulties.

To address concerns about the mispriced bank safety net and potential losses to the insurance fund, Wall (1989) introduces a sub-debt plan aimed at creating a banking environment that, while maintaining deposit insurance, would function like an environment that did not have deposit insurance. Wall’s plan is to have banks issue and maintain “puttable” sub-debt equal to 4 percent to 5 percent of risk-weighted assets. If debtholders exercised the put option, that is, if they required the bank to redeem its debt, the bank would have 90 days to make the necessary adjustments to ensure the minimum regulatory requirements were still satisfied. That is, either retire the debt and continue to meet the regulatory requirement because of excess debt holdings, issue new puttable debt, or shrink assets to satisfy the requirement. If the bank could not satisfy the requirement after 90 days, it would be resolved. The put characteristic has advantages in that it would force the bank to continually satisfy the market of its soundness. Additionally, while earlier plans discussed the need for bond covenants to protect debtholders, all contingencies would be covered under this plan as the market could demand redemption of the bonds without cause. This would essentially eliminate the practice of regulatory forbearance, which was a significant concern at the time, and would subject the bank to increased market discipline. Wall also stresses the need for restrictions on debtholders to limit insider holdings.

Calomiris (1997, 1998, 1999) augments previous sub-debt proposals by requiring a minimal requirement (say 2 percent of total assets) and imposing a yield ceiling (say 50 basis points above the riskless rate). The spread ceiling is seen as a simple means of implementing regulatory discipline for banks. If banks cannot roll over the debt at the mandated spread, they would be required to shrink their risk-weighted assets to stay compliant. Debt would have a two-year maturity with issues being staggered to have equal portions come due each month. This would limit the maximum required monthly asset reduction to approximately 4 percent of assets. To ensure adequate discipline, Calomiris also incorporates restrictions on who would be eligible to hold the debt.10

The effectiveness of any sub-debt requirement depends critically on the structure and characteristics of the program. Most importantly, the characteristics should be consistent with the regulatory objectives, such as increasing direct discipline to alter risk behavior, increasing derived discipline, or limiting or eliminating regulatory forbearance. Keehn, for example, is particularly interested in derived discipline. Wall’s proposal is most effective at addressing regulatory forbearance. Calomiris’s spread ceiling most directly uses derived discipline to force the bank into behavioral changes when the spread begins to bind.

We believe that sub-debt’s greatest value in the near term is as a risk signal. The earliest proposals had limited discussion of the use of sub-debt for derived regulatory discipline. The next round of plans, such as those by Keehn and Wall, use derived discipline, but the only signal they obtain from the sub-debt market is the bank’s ability to issue the debt. We have considerable sympathy for this approach. These types of plans maximize the scope for the free market to allocate resources by imposing minimal restrictions while eliminating forbearance and protecting the deposit insurance fund. However, the cost of providing bank managers with this much freedom is to delay regulatory intervention until a bank is deemed by the markets to be “too risky to save.” As Benston and Kaufman (1988) argue, proposals to delay regulatory intervention until closure may be time inconsistent in that such abrupt action may be perceived by regulators as suboptimal when the tripwire is triggered. Moreover, market discipline will be eroded to the extent that market participants do not believe the plan will be enforced. Benston and Kaufman argue that a plan of gradually stricter regulatory intervention as a bank’s financial condition worsens may be more credible. A version of that proposal, prompt corrective action, was adopted as part of the FDIC Improvement Act of 1991 (FDICIA).
Using sub-debt rates, Calomiris provides a mechanism for this progressive discipline that in theory could last approximately two years. In practice, however, his plan would likely provide the same sort of abrupt discipline as the prior proposals, with the primary difference being that Calomiris’s plan would likely trigger the discipline while the bank was in a stronger condition. His plan requires banks to shrink if they cannot issue subordinated debt at a sufficiently small premium. This would provide a period during which the bank could respond by issuing new equity. If the bank could not or did not issue equity, then it would most likely call in maturing loans to good borrowers and sell its most liquid assets to minimize its losses. However, the most liquid assets are also likely to be among the lowest risk assets, implying that with each monthly decline in size, the bank would be left with a less liquid and more risky portfolio. This trend is likely to reduce most banks’ viability significantly within, at most, a few months. Yet, the previous proposals that would rely on a bank’s ability to issue subordinated debt at any price also give managers some time to issue new equity either by automatically imposing a stay (Wall’s proposal) or by requiring relatively infrequent rollovers (Keehn’s proposal). Thus, Calomiris’s proposal is subject to the same sorts of concerns that arise with the other proposals.

Although Calomiris’s proposal for relying on progressive discipline is more abrupt than it appears at first glance, his suggestion that regulators use the rates on sub-debt provides a mechanism for phasing in stricter discipline. In the next section, we describe our proposal, which offers a combination of Calomiris’s idea of using market rates with Benston and Kaufman’s proposal for progressively increasing discipline.

Our sub-debt proposal differs from previous ones in that it is more comprehensive, with an implementation schedule and a discussion of the necessary changes from current regulatory arrangements. The timing for such reform also seems particularly good as there is a growing consensus that a market-driven means to augment supervisory discipline is needed. Furthermore, banks as a group are relatively healthy, creating an environment in which a carefully thought-out plan can be implemented instead of the hurriedly imposed regulations that sometimes follow a financial crisis.

**A new comprehensive sub-debt proposal**

As discussed earlier, banking organizations’ entry into new activities is raising additional questions about how best to regulate their risk behavior. Ideally, the new activities would avoid either greatly extending the safety net beyond its current reach or requiring costly additional supervision procedures. A plan incorporating sub-debt could help in meeting these challenges. Markets already provide most of the discipline on nondepository financial institutions, as well as virtually all nonfinancial firms. A carefully crafted plan may be able to tap similar market discipline for financial firms to help limit the safety net without extending costly supervision.

Below, we describe our detailed sub-debt proposal. Although our target is the U.S. banking sector, the plan has broader implications as international capital standards come into play. While others have argued that U.S. banking agencies could go forward without international cooperation, we think there are benefits from working with the international banking agencies, if possible. The explicit goals of the proposal are to: 1) limit the safety net exposure to loss, 2) establish risk measures that accurately assess the risks undertaken by banks, especially those that are part of large, complex financial organizations, and 3) provide supervisors with the ability to manage (but not prevent) the exit of failing organizations. The use of sub-debt can help achieve these goals by imposing some direct discipline on banks, providing more accurate risk measures, and providing the appropriate signals for derived discipline and, ultimately, failure resolution.

**Setting the ground rules**

As a starting point, we need to consider whether a new sub-debt program should fit within the existing regulatory framework or require adjustments to the framework in order to effectively fulfill its role. In our view, the goals of the proposal cannot be effectively achieved in the current regulatory environment, which allows banks to hold sub-debt, but does not require that they do so. As a result, banks are most likely to opt out of rolling over maturing debt or introducing new issues precisely in those situations when sub-debt would restrict their behavior and signal the market and regulators that the bank is financially weak. Only a mandatory requirement would achieve the expected benefits. Thus, our proposal requires banks to hold minimum levels of sub-debt.

Similarly, other restrictions in the current regulatory environment limit the potential effectiveness of a sub-debt program. In the current regulatory environment, the role of sub-debt in the bank capital structure is determined by the Basel Accord, which counts sub-debt as an element of tier 2 capital, with the associated restrictions, and limits the amount that may be counted as regulatory capital.
Maintaining the current restrictions has two bothersome implications. First, it dictates almost all of the terms of the sub-debt proposal. For example, U.S. banks operating under current Basel constraints have generally chosen to issue ten-year sub-debt. If there are perceived benefits from having a homogeneous debt instrument, in the current regulatory environment the optimal maturity would appear to be ten years. This is not to say that if left unconstrained financial firms would prefer ten-year maturities. Indeed bankers frequently criticize the restrictions imposed on sub-debt issues that, as discussed above, make it a less attractive form of capital. Ideally, without the restrictions imposed by Basel, the maturity would be much shorter to allow it to better match the duration of the bank balance sheet. However once the ten-year maturity is decided upon as a result of the restrictions, the frequency of issuance is operationally limited to avoid “chopping” the debt requirement too finely. For example, with a 2 percent sub-debt requirement, mandating issuance twice a year would require a $50 billion bank to regularly come to the market with $50 million issues—significantly smaller than standard issues in today’s markets. Thus, adhering to the current Basel restrictions would determine one of the interdependent parameters and thus drive them all. Adjusting the Basel restrictions frees up the parameters of any new sub-debt proposal.

The second implication of following the current Basel Accord is that sub-debt is not designed to enhance market discipline. Given that sub-debt is considered an equity substitute in the capital structure, it is designed to function much like equity and to provide supervisory flexibility in dealing with distressed institutions. In particular, the value of the sub-debt is amortized over a five-year period to encourage banks to use longer-term debt. Furthermore, the interest rate on the debt does not float, thus it is limited in its ability to impose direct discipline when there are changes in the bank’s risk exposure. Finally, because sub-debt is regarded as an inferior form of equity, the amount of sub-debt is limited in the accord to 50 percent of the bank’s tier 1 capital.\footnote{13}

If indeed there are benefits to giving sub-debt a larger role in the bank capital structure, then consideration should be given to eliminating the current disadvantages to using this instrument as capital. That is the approach we take in our proposal.

\textbf{The proposal}

Our sub-debt program would be implemented in stages as conditions permit.

\textbf{Stage 1: Surveillance stage (for immediate implementation)}

- Sub-debt prices and other information would be used in monitoring the financial condition of the 25 largest banks and bank holding companies in the U.S.\footnote{14} Procedures would be implemented for acquiring the best possible pricing data on a frequent basis for these institutions, with supplementary data being collected for other issuing banks and bank holding companies. Supervisory staff would gain experience in evaluating how bank soundness relates to debt prices, spreads, etc., and how changes in these elements correlate with firm soundness.

- Simultaneously, in line with the mandate of the Gramm-Leach-Bliley Act, staffs of regulatory agencies would complete a study of the value of information derived from debt prices and quantities in determining bank soundness and evaluate the usefulness of sub-debt in increasing market discipline in banking. Efforts would be made to obtain information on the depth and liquidity of debt issues, including the issues of smaller firms.\footnote{15}

- If deemed necessary, the regulatory agencies would obtain the necessary authority (via congressional action or regulatory mandate) to require banks and bank holding companies to issue a minimum amount of sub-debt with prescribed characteristics and to use the debt levels and prices in implementing prompt corrective action. The legislation would explicitly prohibit the FDIC from absorbing losses for sub-debt-holders, thus excluding sub-debt from the systemic risk exception in FDICIA.

- The bank regulatory agencies would work to alter the Basel Accord to eliminate the unfavorable characteristics of sub-debt (the 50 percent of tier 1 limitation and the required amortization).

\textbf{Stage 2: Introductory stage (to be implemented when authority to mandate sub-debt is obtained)}

- The 25 largest banks would be required to issue a minimum of 2 percent of risk-weighted assets in sub-debt on an annual basis with qualifying issues at least three months apart to avoid long periods between issues or “bunching” of issues during particularly tranquil times.\footnote{16}

- The sub-debt would have to be issued to independent third parties and be tradable in the secondary market. The sub-debt’s lead underwriter and market makers could not be institutions affiliated with the issuing bank, nor could the debt be held by affiliates. Additionally, no form of credit enhancement could be used to support the debt.\footnote{17}
The terms of the debt would need to explicitly state and emphasize its junior status and that the holder would not have access to a “rescue” under the too-big-to-fail systemic risk clause. It is imperative that the debtholders behave as junior creditors.

Failure to comply with the issuance requirement would trigger a presumption that the bank is critically undercapitalized. If the bank’s outstanding sub-debt trades at yields comparable to those of firms with a below investment grade rating (Ba or lower—that is, junk bonds) for a period of two weeks or longer, then the bank would be presumed to be severely undercapitalized.\textsuperscript{18}

Regulators would investigate whether the remaining capital triggers or tripwires associated with prompt corrective action could be augmented with sub-debt rate-based triggers. The analysis would consider both the form of the trigger mechanism (for example, rate spreads over risk-free bonds or relative to certain rating classes) and the exact rates/spreads that should serve as triggers.

The sub-debt requirement would be phased in over a transition period.

**Stage 3: Mature stage** (to be implemented when adjustments to the Basel Accord allow for sufficient flexibility in setting the program parameters, or at such time as it becomes clear that adequate modifications in the international capital agreement are not possible)

A minimum sub-debt requirement of at least 3 percent of risk-weighted assets would apply to the largest 25 banks, with the expressed intent to extend the requirement to additional banks unless the regulators’ analysis of sub-debt markets finds evidence that the costs of issuance by additional banks would be prohibitive. The purpose is to allow for an increase in the number of banks that can cost effectively be included in the program.

The sub-debt must be five-year, noncallable, fixed rate debt.

There must be a minimum of two issues a year and the two qualifying issues must be at least two months apart.

**Discussion of the proposal**

Stage 1 is essentially a surveillance and preparatory stage. It is necessary because the rest of our proposal requires that regulators have the ability to require sub-debt issuance and access to data to implement the remaining portion of the plan.

At stage 2, regulators introduce the sub-debt program and begin using sub-debt as a supplement to the current capital tripwires under prompt corrective action. The ultimate goal of stage 2 is to use sub-debt-based risk measures to augment capital-based measures, assuming a satisfactory resolution of some practical problems discussed below. The sub-debt tripwires initially set out in stage 2 may reasonably be considered “loose.” Banks that cannot issue sub-debt are probably at or near the brink of insolvency, especially given that they only need to find one issuance window during the course of a year. If a bank’s sub-debt is trading at yields comparable to those of junk bonds, then it is most likely having significant difficulties, and supervisors should be actively involved with the bank. We would not ordinarily expect supervisors to need assistance in identifying banks experiencing this degree of financial distress. However, the presence of such tripwires would reinforce the current mandate of prompt corrective action. Further, it would strengthen derived discipline by other market participants by setting lower bounds on acceptable sub-debt rates.

The use of sub-debt yields for all of the tripwires under prompt corrective action could offer significant advantages. As discussed earlier, market-based tripwires are expected to be more closely associated with bank risk. However, two dimensions need further work before heavy reliance on sub-debt spreads is possible. First, regulators need to review the history of sub-debt rates to determine how best to use them as risk measures and how best to deal with periods of illiquidity in the bond market.\textsuperscript{19} Second, the linking of sub-debt rates to prompt corrective action will imply a tighter link between the prompt corrective action categories and the risk of failure than is possible under the Basel Accord risk measures. Senior policymakers will need to decide where to set the tripwires. What risk of failure is acceptable for a bank to be considered “well capitalized,” “adequately capitalized,” or “undercapitalized”? Thus, at this stage we recommend further study by regulators, academics, and bankers to determine the proper course.

At stage 3, the mature stage, the increased amount of required sub-debt and the shorter maturity should significantly enhance the opportunity for sub-debt to exercise direct market discipline on banks. Another advantage of this proposal is that banks would be somewhat compensated, via the increased attractiveness of sub-debt as regulatory capital, for any increased regulatory burden from holding the additional debt. The removal of the restrictions would make the cost of holding the debt less burdensome than under current regulatory arrangements. While it is not certain, it seems likely that the net regulatory burden would also be less. The five-year maturities in this
appear unnecessary and overly burdensome. They only
debt for overseeing bank activities, the restrictions
However, once the decision is made to employ sub-
restrictions imposed by the Basel Accord.
controversial is the movement toward eliminating the
market expertise.
be an ongoing process as regulators obtain additional
rective action for the categories of well capitalized,
avantageous, because the maturity would more closely
as indicated above, this recalibration will most likely
be an ongoing process as regulators obtain additional
market expertise.
One aspect of our proposal that may appear to be
controversial is the movement toward eliminating the
sub-debt restrictions imposed by the Basel Accord.
However, once the decision is made to employ sub-
debt for overseeing bank activities, the restrictions
appear unnecessary and overly burdensome. They only
serve to increase the cost to participating banks and
to limit the flexibility of the program. Without the
current restrictions, banks would prefer to issue
shorter-term debt and, in some situations, would be
able to count more sub-debt as regulatory capital.
Similarly, as discussed above, the parameters of any
sub-debt policy will be driven in great part by current
regulatory restrictions. Keeping those restrictions in
place would therefore place an unnecessary burden
on participating banks, and would limit regulators,
without any obvious positive payoff. The effort to
adjust Basel also does not slow the movement toward
implementation of a sub-debt program since it would
be phased in through the three-stage process. However,
laying out the broad parameters of the complete plan
in advance would indicate a commitment by regulators
and could increase the credibility of the program.
Once fully implemented, sub-debt would become an
integral part of the regulatory structure.

Concerns and frequently asked questions
about sub-debt

There are a number of issues raised about the
viability of sub-debt proposals. Below, we address
some of these issues and clarify exactly what we
expect sub-debt programs to accomplish. We also
highlight where our proposed sub-debt program spe-
cifically addresses these issues.

Won’t the regulatory agencies “bail out” troubled
institutions by making sub-debt holders at failed in-
stitutions whole if they would have suffered losses
otherwise, thus eliminating the purported benefits of
a sub-debt program? This is probably the most fund-
damental concern raised about the viability of sub-debt
proposals. An implicit guarantee may at times be
more distorting to market behavior than an explicit
guarantee. If debtholders believe that regulators will
make them whole if the issuing bank encounters dif-
culties and cannot make payment on their debt, then
they will behave accordingly. Acting as if they are
not subject to losses, they will fail to impose the nec-
essary discipline on which the benefits of sub-debt
proposals rely. There was evidence of such indiffer-
ence to bank risk levels in the 1980s when the bailout
of the Continental Illinois National Bank ingrained the
too-big-to-fail doctrine into bank investors’ decision-
making. In essence, if the market discipline is not
allowed to work, it will not. This applies to sub-debt.

However, a sub-debt bailout is unlikely under
current arrangements and our proposal makes it even
less likely. Holders of sub-debt are sophisticated inves-
tors, who understand their position of junior priority
and the resulting potential losses should the issuing
firm encounter difficulties. Additionally, since banks
are not subject to bankruptcy laws, debtholders can-
ot argue for a preferred position by refusing to accept
the bankruptcy reorganization plan. Thus, they are
unable to block the resolution. So pressures to rescue
debtholders should not arise either from a perceived
status as unsophisticated investors or from their bar-
gaining power in the failure resolution process.

The FDIC guaranteed the sub-debt of Continental
of Illinois in 1984, but it did so to avoid having to
close the bank and not to protect the sub-debt investors
per se. The effect of FDICIA and its prompt correc-
tive action, least cost resolution requirements, and
too-big-to-fail policies was to significantly curtail
and limit the instances when uninsured liability holders
would be protected from losses. Benston and Kaufman
(1998) find that policy did change as a result of
FDICIA, as significantly fewer uninsured depositors
were protected from losses at both large and small
banks after passage of the legislation. Similarly,
Flannery and Sorescu (1996) find evidence that the
markets viewed FDICIA as a credible change in policy
and, as a result, sub-debt prices began reflecting dif-
cferences in bank risk exposures. Thus, the market
apparently already believes that sub-debt-holders
will not be bailed out in the future.

Under our sub-debt proposal, there would be
still lower potential for debtholder rescue. Unlike
depositors, who can claim their assets on demand, holders of the intermediate-term debt could only claim their assets as the debt matured instead of initiating a “bank run,” the kind of event that has typically prompted the rescues we have seen in the past. Additionally, there is much less subjectivity if the sub-debt price spreads are used for prompt corrective action rather than book value capital ratios. Finally, under our proposal, the sub-debt holder would be explicitly excluded from the class of liabilities that could be covered under the systemic risk exception. This exclusion should be viewed favorably by banks. Under the terms of the too-big-to-fail exception in FDICIA, losses from the rescue would have to be funded via a special assessment of banks. Therefore, banks should encourage the FDIC to strictly limit the extent of the liabilities rescued.

*Are there cost implications for banks?* Interestingly, the costs associated with issuing sub-debt have been used as an argument both for and against sub-debt proposals. The standard argument is that there are relative cost advantages from issuing debt resulting from the favorable tax treatment. It is also argued that closely held banks may find debt to be a less expensive capital source as new equity injections would come from investors who realize they will have a minor ownership role. Both arguments suggest that an increased reliance on sub-debt would result in cost savings.

There are, however, some additional actual or potential costs associated with increased sub-debt issues. First, increased reliance on relatively frequent debt rollovers would generate transaction costs or issuance costs. There is disagreement as to just how expensive these costs would be. Some argue that the cost would be similar to that required for issuing bank certificates of deposit, while others argue that the cost could be quite substantial. The issuance frequency discussed in most sub-debt proposals, however, is not very different from the current frequency at large banking organizations. Two issues per year, which is well within the recommendations in most sub-debt proposals, is relatively common in today’s banking markets.

A more significant concern seems to be where, within the overall banking organization, the debt would be issued. Most sub-debt proposals require the debt to be issued at the bank level whereas, until recently, most sub-debt was issued at the bank holding company level. This allowed the holding company the flexibility to distribute the proceeds throughout the affiliated firms in the organization. This occurred in spite of the fact that the rating agencies typically rated bank debt higher than the debt of the holding company, and, similarly, holding company debt typically traded at a premium to comparable bank debt. This would suggest that the additional flexibility from issuing debt at the holding company level is of value to the banking organization. Removal of this flexibility would impose costs. The recent trend toward issuing more debt at the bank level, however, would suggest the value of this flexibility is becoming less important.

A more important cost implication is imbedded in our sub-debt proposal. In the past, regulators have restricted the use of sub-debt by limiting the amount that could count as capital and by requiring that the value of the sub-debt be amortized over the last five years before maturity. These restrictions are imposed because the firm needs to make periodic payments on the debt, regardless of its financial condition. However, this does not decrease the effectiveness of sub-debt in serving the capital role as a cushion against losses. It still buffers the insurance fund. By eliminating these restrictions in our sub-debt proposal, we enhance the value of the debt as capital and decrease the net cost of introducing the proposal.

*Isn’t there a problem in that sub-debt proposals are procyclical?* A possible concern with sub-debt requirements is that they may exacerbate procyclical behavior by banks—increased lending during economic expansions and reduced lending during recessions. However, this is not unique to sub-debt programs; any regulatory requirement that does not adjust over the course of a business cycle has the potential to be procyclical if banks seek to only satisfy the minimum requirements. For example, appendix D of Kwast et al. (1999) points out that bank capital adequacy ratios are likely to decline during recessions as banks experience higher loan losses, implying that regulation based on capital adequacy ratios has the potential to be procyclical.

The procyclicality of a regulatory requirement may be at least partially offset if banks seek to maintain some cushion above minimum regulatory requirements that they may draw on during economic downturns. In the case of the regulatory capital adequacy requirements, both casual observation of recent bank behavior and formal empirical analysis from the 1980s and early 1990s suggest that banks do indeed seek to maintain such a cushion for contingencies.

Moreover, a regulatory program that uses sub-debt yields as triggers for regulatory action may be designed to induce less procyclical behavior than would other types of regulatory requirements. Consider two ways to design the sub-debt triggers as discussed in Kwast...
et al. (1999). One design is to base regulatory action on a constant basis point spread over bonds with little or no credit risk, such as Treasury securities. Such a standard is more likely to become binding during recessions when banks are experiencing loan losses and investors demand higher risk premiums to continue holding bank bonds. Thus, a policy that sets triggers at a constant premium over Treasuries may result in procyclical regulation in a manner similar to that of standard capital requirements.

Another way of designing the triggers, however, is to base them on a measure that offers countercyclical yields over the business cycle, for example, the yields on corporate bonds of a given rating. There is evidence that bond-rating agencies seek to smooth ratings through business cycles. For example, Theodore (1999, p. 10) states Moody’s policies:

Moody’s bank ratings ... aim at looking to the medium- to long-term, through cyclical trends. For example, a drop in quarterly, semi-annual or even annual earnings is not necessarily a reason to downgrade a bank’s ratings. However, if the earnings drop is the result of a structural degradation of a bank’s fundamentals, credit ratings need to reflect the new developing condition of the bank.

If the rating agencies are trying to “look through the business cycle,” then the spreads on corporate bonds over default-free securities should be small during expansions because investors, but not the rating agencies, recognize a lower probability of default during expansions. Similarly, the spreads on corporate bonds over default-free bonds should rise during recessions as the markets, but not the rating agencies, recognize the increased probability of default. Thus, prompt corrective action triggers based on sub-debt yields relative to corporate yields introduce an element of smoothing. The triggers may be relatively tight during expansions when banks should be building financial strength and relatively loose during downturns as they draw down part of their reserves.

One case where the use of sub-debt yields may tend to reinforce the business cycle is when liquidity drops in all corporate bond markets and risk premiums (including liquidity risk premiums) temporarily soar. However, our proposal recognizes this potential problem and provides for temporary relief until liquidity improves.

Aren’t supervisors better gauges of the riskiness of a bank because they know more about each bank’s exposure than the market does? If so, then why not rely exclusively on the supervisors instead of holders of sub-debt? In some cases the market’s knowledge of a bank’s exposure may indeed be a subset of the examiner’s knowledge. However, we rely on markets to discipline firm risk taking in virtually every other sector of our economy, so markets must have some offsetting advantages. One such advantage is that the financial markets are likely to be better able to price the risks they observe because market prices reflect the consensus of many observers investing their own funds. Another advantage of markets is that they can avoid limitations inherent in any type of government supervision. Supervisors are rightfully reluctant to be making fundamental business decisions for banks unless or until results confirm the bank is becoming unsafe or unsound. Further, even when supervisors recognize a serious potential problem, they have the burden of being able to prove to a court that a bank is engaged in unsafe activities. In contrast, in financial markets the burden of proof is on the bank to show it is being safely managed. A further weakness of relying solely on bank supervisors is that they are ultimately accountable to the political system, which suggests that noneconomic factors may enter into major decisions no matter how hard supervisors try to focus solely on the economics of a bank’s position. Sub-debt investors have no such accountability; they may be expected to focus solely on the economic condition of individual banks.

A typical concern surrounding sub-debt proposals is that the perceived intent is to supplant supervisors and rely solely on the forces of the marketplace to oversee bank behavior. In our proposal, the intent is to augment, not reduce supervisory oversight. If supervisors have additional information about the condition of a bank, there is nothing in the sub-debt proposals limiting their ability to impose sanctions on the activities of the bank. In addition to sub-debt serving the standard role as a loss-absorbing capital cushion, it serves as an additional tool for use by both the private markets and the regulators to discipline banks objectively. In fact, one of the major components of our proposal is to have the supervisors incorporate the yield spreads for use in prompt corrective action. With private markets providing information, supervisors can focus their efforts on exceptional circumstances, leaving the well-understood risks for assessment by the marketplace.

Do we currently know enough about the sub-debt market to proceed? Although we would like to know more about the sub-debt market, we think considerable information is already available. The studies surveyed and the new evidence presented in Kwast et al. (1999) provide considerable insight into the subordinated debt market. These studies suggest that investors in sub-debt do discriminate on the basis of the riskiness of their portfolios.
Moreover, a review of the regulatory alternatives suggests that any durable solution to achieving an objective measure of banks’ risk exposure will look something like our proposal. The problems that plague the existing risk-based capital guidelines are inherent in any attempt by the supervisors to measure the riskiness of a bank’s portfolio based on a presupposed set of criteria. Over time, banks will find or will manufacture claims whose intrinsic contribution to the riskiness of the bank’s portfolio is underestimated by the supervisory criteria. That is, banks will attempt to arbitrage the capital requirements.

An alternative to supervisory determined criteria is to use market evaluations. The Basel Committee on Banking Supervision correctly moved in this direction with its proposed new capital adequacy framework. However, it chose to ask opinions of market participants rather than observing market prices and quantities. The committee then compounded this by proposing to ask the opinions of the two parties, the banks and their rating agencies, that have incentives to underestimate the true risk exposure.

A superior system for obtaining a market-based risk measure will use observed data from financial markets on price or quantity, or both. That is, it will use a market test. The relevant question to be addressed is which instruments should be observed, how these instruments should be structured, and how supervisors can best extract the risk signal from the noise generated by other factors that may influence observed prices and quantities. In principle, any uninsured bank obligation can provide the necessary information. We favor sub-debt because we think it will provide the cleanest signal.

There are alternatives to sub-debt. Common equity may currently have the advantages of being issued by all large banks and of trading in more liquid markets. However, investors in bank common equity will sometimes bid up stock prices in response to greater risk taking, so their signal can only be interpreted in the context of a model that removes the option value of putting the bank back to the firm’s creditors (including the deposit insurer). In contrast, valuable information can be extracted from subordinated debt without a complicated model. If a bank’s debt trades at prices equivalent to Baa corporate bonds, then its other liabilities are at least Baa quality.

Banks also issue a variety of other debt obligations that could be used to measure their risk exposure. The use of any debt obligation that is explicitly excluded from the systemic risk exception in FDICIA could provide a superior risk measure to those proposed by the Basel Committee. Thus, we conclude that sub-debt is the best choice because it is the least senior of all debt obligations if a bank should fail and, therefore, its yields provide the clearest signal about the potential risk that the bank will fail. We think sufficient information exists to adopt a sub-debt proposal with the understanding that the plan will be refined and made more effective as additional information and analysis become available.

Conclusion

FDICIA sought to reform the incentives of both banks and their supervisors. The least cost resolution provisions were intended to expose banks to greater market discipline and the prompt corrective action provisions were intended to promote earlier and more consistent supervisory discipline. Ongoing developments are undercutting both sources of discipline. Whether the government would have been willing to take the perceived short-term risks associated with least cost resolution procedures for a very large bank immediately after their introduction is debatable. Arguably, those risks have increased significantly as banks have grown larger and more complex.

Whether prompt corrective action based on book values would have been effective in closing banks before they became economically insolvent is also questionable. Unquestionably, however, banks’ ability to “game” regulatory risk measures has grown over the last decade.

Although ongoing developments are undercutting the intent of FDICIA, the premise that banks and their supervisors should be subject to credible discipline remains. Ideally, this discipline would come from financial markets. While markets do not have perfect foresight, they are both flexible enough to accept promising innovations and willing to acknowledge their mistakes, even if such recognition is politically inconvenient.

Sub-debt provides a viable mechanism for providing such market discipline. It is already providing useful signals in today’s financial markets. We propose to combine these signals with the gradual discipline provided under prompt corrective action in a form that is credible to banks and other financial market participants.

This article provides a feasible approach to implementing enhanced discipline through sub-debt. Our proposal draws on the existing evidence on market discipline in banking and the insights of previous proposals and policy changes. The new plan provides for phased implementation and leaves room for future modifications as additional details concerning the market for sub-debt are determined. The plan calls
for specific changes in those areas where we believe the evidence is relatively clear, such as the fact that large solvent banks should be able to issue sub-debt at least once a year. In those areas where the evidence is weak to non-existent, we defer decisions pending additional study. This approach should enhance the credibility of the plan. Although the details of the plan would evolve over time, once the basics are implemented the industry and the public would see bank behavior being significantly influenced by both market and supervisory oversight. The combination should make for a more effective, safe, and sound industry.

NOTES

1See Title 1, Section 108 of the Gramm-Leach-Bliley Act entitled “The use of subordinated debt to protect the deposit system and deposit system funds from ‘too big to fail’ institutions.”

2During crises, the pressure of having to respond quickly increases the likelihood of introducing poorly structured regulation. Industries where regulatory reforms introduced during crises may have caused significant long-term problems include banking in the 1930s (Kaufman, 1994) and the pharmaceutical industry following the infamous Thalidomide incidents in the 1950s (Evanoff, 1989).

3An index of papers that can be downloaded from the Basel Committee on Banking Supervision website may be found at www.bis.org/publ/index.htm.

4See Bank for International Settlement (1999).

5The rating agency obviously has an incentive to maintain its credibility as an objective entity and could resist the pressure. The incentives, however, would work in this direction.

6More generally, in recent years there has been growing concern about the need to increase the role of market discipline in banking. See, for example, Ferguson (1999), Meyer (1999), Stern (1998), Boyd and Rolnick (1998), Broadus (1999), and Moskow (1998).

7Direct discipline would result from an expected increase in the cost of issuing debt to response to an increase in the bank’s perceived risk profile. To avoid this increased cost the bank would more prudently manage risk. Derived discipline results when other agents (for example, supervisors) use the information from sub-debt markets to increase the cost to the bank. For example, as discussed below, bank supervisors could use debt yields as triggers for regulatory actions.

8Additional discussion of the role of sub-debt in this plan can be found in Evanoff (1993, 1994).

9Regulatory restrictions would be prompt-corrective-action-type constraints such as limits to dividend payments or deposit and asset growth rates once core equity fell below 2 percent of risk-weighted assets.

10The sub-debt requirement is one component of Calomiris’s regulatory reform proposal aimed at modifying industry structure and the operating procedures of the International Monetary Fund. It would also include a mandatory minimum reserve requirement (20 percent of bank debt in Calomiris, 1998), minimum securities requirement, and explicit deposit insurance. Although some details of his proposal, such as requiring the debt be issued to foreign banks, may not be feasible for U.S. banks, the general approach provides interesting insights into the issues in designing a sub-debt plan for the U.S.

11This is not the first time proposals have suggested sub-debt be linked with prompt corrective action; see Evanoff (1993, 1994) and Litan (2000).

12The term banking is used generically and could include all depository institutions.

13As discussed earlier, the current bank capital requirement framework is being reevaluated (see Bank for International Settlements, 1999). As part of the debate, some have recommended total elimination of the tier 1 versus tier 2 distinction, (for example, Litan, 2000). If this approach is taken, we would recommend that minimum leverage requirements be maintained to ensure sufficient levels of equity (although it would be in sub-debt holders self interest to ensure this occurs) and to provide supervisors with an official tool for intervening when equity levels fall to unacceptable levels.

14When fully implemented, the policy would apply to “banks” instead of the bank holding company. During this surveillance stage, however, information could be gained at both levels.

15Actually, progress is currently being made on these first two items. The Board staff are actively involved in collecting and analyzing sub-debt price data, and System staff are evaluating how the markets react to debt spreads.

16The only exception would occur if general market conditions precluded debt issuance by the corporate sector (both financial and nonfinancial firms). This exception requires more specific details, but it would be an industry-wide rather than a bank-specific exception.

17The objective is to limit “regulatory gaming”; see Jones (2000). Additional minimum denomination constraints could be imposed to further ensure that debtholders are sophisticated investors. (for example, see U.S. Shadow Financial Regulatory Committee, 2000).

18Depending on the depth of the secondary market, this may need to be extended to a couple of weeks. Again, the timeframe could be modified as more market information is obtained. Additionally, to allow for flexibility under extreme conditions, procedures could be introduced by which the presumption could be overturned given the approval of the FDIC upon request by the bank’s primary federal supervisor. The procedures for this exception, however, would be somewhat similar to those currently in place for too-big-to-fail exceptions, for example, submission of a public document to Congress, et al.

19For example, should risk be measured as the spread between the yield on a sub-debt issue and a comparable maturity Treasury security, the yield on a bank’s sub-debt versus the yield on comparable maturity corporate bonds in different ratings classes, or the spread over LIBOR (London Interbank Offered Rate) after the bond is swapped into floating rate funds.
This is not to say that initiating changes to the accord would be costless. Obviously negotiations would be required since other country members may want to continue to have sub-debt be an inferior form of capital. But from the participating U.S. banks’ perspective and the regulators’ perspective (concerning program flexibility), the elimination of these restrictions should result in net benefits.

We are not saying that detailed parameters should be introduced at this time. As argued above, additional analysis is required before these could be decided upon.

Another potential issue is how the banks will respond to the new regulation in an attempt to avoid sub-debt discipline. A review of this issue is included in Kwast et al. (1999), and our proposal raises no new concerns. The recently passed Financial Services Modernization Act addresses some of these potential concerns by significantly limiting credit enhancements on sub-debt.

Jones (1998) suggests the cost of equity could be twice that of debt once the tax differences are accounted for. Benston (1992) discusses the cost differences and other advantages of sub-debt over equity capital.

Alternatively, the current owners could inject equity but that may be costly in that it places them in a situation where they are relatively undiversified.

For example, see Kwast et al. (1999). The exception is Calomiris (1998) which would require monthly changes via either debt issues or asset shrinkage.

This holding company premium is typically associated with the bank having access to the safety net and the associated lower risk of default during times of financial stress. Alternatively, it has been argued the differential results from the different standing of the two debtholders. Holders of bank debt have a higher priority claim on the assets during liquidation of the bank than do the holders of holding company debt which essentially has an equity claim on the bank.

The liquidity crunch in the fall of 1998 and the Long-Term Capital episode are possible examples of such a problem period.

For example, the American Banker reports that the Office of the Comptroller of the Currency is threatening to downgrade bank’s safety and soundness rating if they fail to supply accurate Community Reinvestment Act data; see Seiberg (1999).

Supervisory agencies could short circuit this avoidance by having their examiners conduct subjective evaluations but that could easily result in examiners serving as shadow managers of banks.

Preferred stock is a form of equity but it would yield a clean signal unlike common equity. We do not propose the use of preferred stock for two reasons. First, dividend payments on preferred stock are not a deductible expense to the bank. Thus, forcing them to issue preferred stock would increase their costs. Second, discussions with market participants, as reported in Kwast et al. (1999, p. 45), indicated that the preferred stock market is more heavily influenced by “relatively uninformed retail investors.”

REFERENCES


Seiberg, Jaret, 1999, “CAMEL’s penalty threatened if flaws found in CRA data,” American Banker, April 27, p. 2.


Unemployment and wage growth: 
Recent cross-state evidence

Daniel Aaronson and Daniel Sullivan

**Introduction and summary**

The current economic expansion, now the longest on record, has delivered the lowest unemployment rates in 30 years. Yet nominal wage growth has remained relatively contained. This failure of wages to accelerate more rapidly suggests to some a shift, or even a complete breakdown, in the historical relationship between unemployment and wage growth. However, looking across the years, the relationship between unemployment and wage growth has always been relatively loose, implying that it might take many years to conclusively identify even a significant change in the link between unemployment and wages.

In this article, we look across the states for more timely evidence of a change in the relationship between unemployment and wage growth. We find, however, that even in recent years, there is a relatively robust, negative relationship between state unemployment rates, properly evaluated, and wage growth. In particular, states in which current unemployment rates are lower relative to their long-run averages tend to have faster wage growth than those in which unemployment is higher relative to average. We do find some evidence that the sensitivity of wage growth to unemployment may have decreased in recent years, but we consider that evidence to be somewhat weak.

Before turning to the cross-state evidence, we briefly review some of the cross-year evidence that has led to speculation about a change in the relationship between unemployment and wage growth. That speculation has taken a number of forms, not all of which have been well reasoned. In particular, media analysts sometimes have characterized the lack of greater acceleration of nominal wages in the face of low unemployment as a failure of the “forces of supply and demand” in the labor market. But, the forces of supply and demand have direct implications not for nominal wage growth, but rather for real, or inflation-adjusted, wage growth. Indeed, because nominal wage growth depends on the level of price inflation, which in turn depends on monetary policy, there is little reason to expect a long-run link between the level of nominal wage growth and unemployment. So it is not surprising that the statistical relationship between nominal wage growth and unemployment discovered by Phillips (1958) disappeared long ago.

A more serious question is whether there has been a change in the relationship between unemployment and the growth of wages relative to expected inflation. A rough indication of the time-series evidence on this question can be gleaned from figures 1 to 3, which are scatter plots of annual data on the excess of wage growth over the previous year’s price inflation versus the natural logarithm of the annual unemployment rate. In each case price inflation is measured by the change in the log of the annual Consumer Price Index. The three figures differ, however, in their measures of wage growth.

In figure 1 wage growth is the change in the log of the annual average of the Bureau of Labor Statistics’ (BLS) Average Hourly Earnings (AHE) series. This closely followed monthly wage measure is limited to the wage and salary earnings of the approximately 80 percent of private industry workers who are classified as production or nonsupervisory workers. In figure 2 wage growth is derived from the hourly compensation measure from the BLS’s productivity and cost data (Hourly Comp). This measure captures most wage and nonwage forms of compensation paid to all workers in the business sector and thus provides a superior measure of the compensation associated
with an average hour of work. Finally, in figure 3 wage growth is given by the increase in the average value of the BLS’s Employment Cost Index (ECI). This measure also reflects both wage and benefits costs for private employers and, in addition, adjusts for variation in the industrial and occupational mix of the labor force. Unfortunately, it only became available in 1983. So there are relatively few observations in figure 3.

The relationships depicted in figures 1–3 are analogous to the wage equations in some macroeconometric models. They can be motivated by assumptions that 1) wages are set to exceed expected inflation by an amount that depends on the unemployment rate, and 2) expected inflation is equal to the level of inflation in the previous year. Of course, wage equations in actual macroeconometric models are considerably more elaborate than what is represented in the figures. In particular, they use quarterly rather than annual data and they allow for more complicated dynamics. They also include other variables, such as the level of productivity, that influence wage growth.

Nevertheless, figures 1–3 illustrate the basic nature of the time-series evidence on the relationship between wages and unemployment.

In at least the first two figures, there is a loose, but reasonably clear, negative correlation between unemployment and wage growth in excess of lagged inflation. The least squares regression lines shown in the figures all slope downward with elasticities that range from –0.044 for AHE to –0.055 for Hourly Comp to –0.013 for the ECI. The estimated standard errors of these estimates are 0.0095, 0.0090, and 0.0090. Thus, if the relationships are stable over time, one can be reasonably confident that the true coefficients are different than zero for AHE and Hourly Comp. For the ECI, the evidence is less clear-cut, in part, perhaps, because the available sample is much shorter. Of course, in all three figures there is a sizable spread of values around the estimated line; the relationship between unemployment and wage growth is far from tight.

The data for the current expansion are highlighted in figures 1–3 by a line connecting the values from 1992 to 1999, when the unemployment rate was falling from 7.5 percent to 4.2 percent. Evidently, the extent of departure of recent data from historical patterns depends a good deal on the measure of wage growth. On the one hand, the recent AHE data shown in figure 1 have stayed remarkably close to the typical pattern. AHE growth from 1992 to 1999 did not differ from the estimated regression line by more than four-tenths...
of a percentage point, while in some earlier years the deviation had been as much as 2 percentage points. On the other hand, the more comprehensive Hourly Comp data shown in figure 2 have departed fairly significantly from expectations over much of this expansion. In particular, the growth of Hourly Comp was a percentage point or more below expectations each year from 1993 to 1997. Though the data for the last two years have returned to the predicted line, the cumulative loss of wage growth over the expansion has been significant. Finally, the recent ECI data shown in figure 3 have also departed rather significantly from historical norms. As with the Hourly Comp data, ECI growth was significantly below expectations early in the expansion. But growth actually exceeded expectations late in the expansion, so the cumulative difference in wage growth is considerably less.

The differences in the performance of the three wage measures reflects the differing pattern of growth in wage and nonwage compensation over the sample periods as well as the coverage of the measures. Over most of the period covered in the graphs, nonwage compensation grew faster than wage compensation. For instance, according to data from the National Income and Product Accounts, the fraction of employee compensation paid in the form of wage and salary accruals fell from 92.4 percent in 1959 to 83.4 percent in 1980 to a minimum of 81.0 percent in 1994. Since 1994, however, the fraction of compensation paid in the form of wages and salaries has increased to 83.9 percent (in 1999), holding the growth of total compensation measures such as Hourly Comp and the ECI below that observed for AHE. In addition, over much of the period covered in the figures, wage growth has been more rapid for the more highly skilled, who are less likely to be classified as production and nonsupervisory workers and thus less likely to be covered in AHE.

Taken together, the evidence in figures 1–3 for a significant recent shift in the relationship between unemployment and expected real wage growth appears to us to be relatively weak. As we have noted, when one focuses on the more comprehensive Hourly Comp measure, the departures from expectations over this expansion have at times been relatively great. But, such departures are far from unprecedented. In earlier years, the data have strayed further from expectations only to return to the basic pattern of low unemployment being associated with higher growth of wages relative to lagged inflation. Of course, the evidence in figures 1–3 also does not rule out a significant shift in the relationship between unemployment and inflation. Unfortunately, given the looseness of the historical relationship, it would take many years to confidently identify even a relatively large change in the relationship.

Some shift in the relationship between unemployment and wage growth would not be terribly surprising. Among the many changes in the labor market in recent years, the general drop in the level of job security, the aging of the work force, its higher levels of education, the growth of temporary services employment, the use of fax machines and the Internet in job search, and even the increase in the prison population could each be changing the relationship between unemployment and wage growth.\(^7\)

Moreover, the theoretical basis for the relationships depicted in the figures is somewhat loose, which at least suggests the possibility of instability. The assumption that expectations of inflation are equal to last year’s level of inflation is clearly ad hoc. Moreover, though a relationship between expected real wage growth and unemployment can be motivated by economic theory, such theory doesn’t necessarily imply a special place for the standard civilian unemployment rate.

Indeed, in the simplest model of a competitive labor market, unemployment is not a well-defined concept because there is no distinction between workers...
being unemployed and out of the labor force. Rather, in that model wages adjust to clear the market, and workers for whom the equilibrium wage is below the alternative value of their time simply choose not to work. The competitive model would replace the relationship in figures 1–3 with a standard, aggregate labor supply curve. This is analogous to the relationship in figures 1–3, but with employment, rather than unemployment, as the variable predicting wage growth. Of course, (deviations from trend) fluctuations in these variables are highly correlated, so unemployment may predict expected real wage growth reasonably well even if employment is the theoretically preferable measure.

Economic theorists have gone beyond the simple competitive framework to formulate models in which unemployment is involuntary and in which the unemployment rate is related to wages. One class of such models explicitly recognizes the importance of the labor market search, the complex process by which workers desiring jobs and firms desiring workers are matched to each other. In such models, some workers and firms are left unmatched and thus unemployed or with vacancies. Moreover, in search models with wage bargaining, workers have greater bargaining power when the unemployment rate is low, since turning down a job offer with a low wage is more palatable when the unemployment rate is low. This generates a link between unemployment and wages.

Another class of models in which unemployment can be involuntary and in which the unemployment rate is connected to wages incorporates what are known as efficiency wage considerations. In such models, involuntary unemployment arises because firms rationally choose to pay wages above market clearing levels in order to induce effort or reduce turnover. For instance, when it is difficult to monitor workers’ effort, firms may want to ensure that workers truly fear being discharged after having been found to exert insufficient effort. This will be the case if wages are high enough that workers prefer working to being unemployed. In such models, wages cannot fall enough to clear the labor market because if they did so, workers would have insufficient incentive to put forth appropriate effort. The connection of wages to unemployment emerges because when unemployment is low, discharged workers will face less time out of a job. Thus, wages need to be further above the value of workers’ nonmarket uses of time to induce the same level of effort.

Even in search and efficiency wage models, the standard unemployment rate may not be the variable most directly related to wages. Rather, in both classes of models, the exit rate, the rate at which workers leave unemployment, is a more direct measure of the cost to workers of becoming or staying unemployed than the unemployment rate itself, which also depends on the rate of entry into unemployment. Of course, since the exit rate and the overall unemployment rate are highly correlated, the latter may predict wages reasonably well even if the former is the variable that is truly linked to expected wage growth.

Even if one accepts the use of an unemployment rate as the measure of labor market conditions, there is still the question of which unemployment rate to use. The standard measure imposes requirements that nonemployed workers be available for work and have made an effort to find work in the last month. However, some out-of-the-labor-force workers, for example, those who say they want a job, are relatively similar to the unemployed and may exert an influence on wage growth. Conversely, some of those who are unemployed, such as those who have been unemployed for long periods, may be more similar to the out-of-the-labor-force pool. Ultimately, which measure best captures the labor market forces influencing wages is an empirical question, the answer to which could be changing over time.

In this article we look for evidence of such changes in the cross-state relationship between unemployment and wage growth. Previous work has demonstrated a relationship between unemployment and wage growth across states that is analogous to that in time-series data. The basic assumption underlying this work is that inflation expectations are approximately the same for all states in a given year. Given that the U.S. has a single, national monetary policy, this is plausible, though clearly one could imagine deviations from this assumption. If inflation expectations are constant across states, differences in wage growth across states are unaffected by inflation expectations. Similarly, to the extent that other variables, such as productivity, that affect wage growth are constant across states in a given year, comparisons of states’ wage growth rates are also unaffected by these variables.

A major advantage of the cross-state approach is the greatly increased number of degrees of freedom available from the wide variation in state unemployment rates. This makes it possible to estimate the response of wage growth to unemployment separately for relatively short periods. Thus, it may be possible to identify changes in that response that would take many years of time-series data to uncover.

Despite its attractions, the cross-state approach requires some care in its implementation. In particular,
differences across states in unemployment rates persist for long periods, reflecting differences in factors such as demographics, industry composition, and generosity of social insurance that don’t necessarily translate into differences in wage growth. The cross-state approach can allow for such persistent differences across states by employing multiple years of data. The empirical analysis then amounts to measuring the tightness of a state’s labor market by its deviation from its own average unemployment rate over the entire sample period.

Deviations from mean unemployment rates reveal a different view of where labor markets are tight than the simple level of unemployment. For example, Wisconsin unemployment averaged 3.1 percent in 1999, six-tenths of a point less than in Michigan where unemployment averaged 3.7 percent. But, Michigan has historically had much higher unemployment than Wisconsin. For instance, over the 1980–99 period, Michigan’s average unemployment rate was 8.4 percent, versus 5.7 percent in Wisconsin. Thus, Michigan in 1999 was 4.7 percentage points below its average, while Wisconsin was only 2.6 points below its average. Our empirical analysis finds that such unemployment-deviation measures are a better guide to labor market tightness than the standard unemployment rate.

That empirical work confirms the negative cross-state correlation between unemployment and wage growth found by previous researchers for the years 1980–99. We also find that the elasticity of wages with respect to unemployment has fallen over successive five-year intervals, a result that does not seem to be the result of a compositional shift toward college-educated workers. However, we regard this evidence of a weakened relationship between unemployment and wage growth as itself somewhat weak. In particular, when we estimate an elasticity for each year from 1980 to 1999, there is enough year-to-year variability that a downward trend in the magnitude is not obvious. Rather, the extent of change observed in the relationship depends on the necessarily arbitrary decision of where to draw the line between periods. Moreover, if one considers the response of wage growth to the level of unemployment, rather than its logarithm, there is very little evidence of a recent change in the sensitivity of wage growth to unemployment.

A recent study by Lehrman and Schmidt (1999) of the Urban Institute for the U.S. Department of Labor suggests that the level of unemployment across states is not related to wage growth. We believe those authors’ results differ from ours for at least the following reasons: their measure of unemployment is not well matched in time to their measure of wage growth, their procedure does not allow for differences across states in other factors that affect wage growth, and their statistical procedure, which does not impose a linear relationship between wage growth and unemployment, has high variability with only 50 state observations. Thus, we agree with Zandi (2000), who concludes that the results of Lehrman and Schmidt (1999) prove little about the relationship between unemployment and wage growth.13

Our main results concern possible changes in the sensitivity of wage growth to unemployment. But we also briefly examine how the level of wage growth for particular levels of unemployment may have changed over time. We find that the levels of real wage growth associated with high, medium, and low unemployment rates have been reasonably constant in recent years. The real wage growth levels associated with typical values of unemployment were somewhat higher in the early 1980s, but since then have been relatively constant, with the wage growth associated with high unemployment rates actually rising somewhat in the late 1990s. Similarly, the unemployment rate associated with the average rate of real wage growth fell after the early 1980s, but has been relatively constant since then.

Because, as we noted, there is no compelling theoretical reason for the standard civilian unemployment rate to be the best measure of labor market conditions for predicting wage growth, we investigated a number of alternative measures of labor market tightness. These included the employment-to-population ratio, broader and narrower measures of unemployment, separate measures of short-term and long-term unemployment, and a measure of the exit rate from unemployment. Most of these measures predict wage growth about as well as the standard unemployment rate. Most also show the same decline in the magnitude of their elasticity with respect to wage growth that we observe over five-year intervals for the unemployment rate. The decline in the coefficients associated with the exit rate and short-term unemployment measures are, however, more severe. Such findings suggest that further work on improved measures of labor market tightness may be fruitful.

Finally, our results have implications for inflation forecasting, a task that plays an important role in the formulation of monetary policy. One of the most widely used approaches to such forecasting has been the short-run, or expectations-augmented, Phillips curve.14 This forecasting method, which relates the change in price inflation to the level of the unemployment rate and other variables, can be derived from
the kind of expected real wage growth relationship depicted in figures 1–3 along with an equation that relates price inflation to wage inflation and other variables.\textsuperscript{15} Recently, there is evidence that typical short-run Phillips curve specifications have systematically overforecasted inflation.\textsuperscript{16} Our results point toward the conclusion that this failure of the forecasts is most likely attributable to the part of the model linking price inflation to wage growth rather than to a change in the relationship between expected real wage growth and unemployment. This is consistent with the findings of Brayton et al. (1999), who show that including additional variables related to the markup of prices over wages helps to stabilize the Phillips curve.

Data

Our main results are based on two data sources. The first is the annual averages of the standard, monthly, state-level unemployment rates reported by the BLS. The second source is a measure of state-level, demographically adjusted wage growth that we construct from the micro data of the outgoing rotations of the Current Population Survey (CPS). The CPS, which is the source for such well-known statistics as the unemployment rate, is a monthly, nationally representative survey of approximately 50,000 households conducted by the Census Bureau.\textsuperscript{17} Households in the CPS are in for four months, out for the following eight months, and then in again for four more months. Those in the fourth and eighth month of their participation are known as the outgoing rotation groups (ORG) and are asked some additional questions, including their earnings in the previous week. We compute an individual’s hourly wage rate as the ratio of weekly earnings to weekly hours of work.\textsuperscript{18} Pooled across the 12 months of the year, the ORGs yield an annual sample size of at least 150,000 households. They are available starting in 1979.

We summarize the individual-level wage data with an adjusted average wage for each state-year pair. These are obtained as state-year-specific intercepts in a regression of the natural logarithm of wages on demographic and educational characteristics:

\begin{equation}
\omega_{ist} = \omega_{i} + x_{ist} \beta + \eta_{ist},
\end{equation}

where $\omega_{ist}$ is the log of the wage for individual $i$ in state $s$ and year $t$. The vector $x_{ist}$ of control characteristics is the same as that utilized by Blanchard and Katz (1997) and consists of a quartic in potential experience interacted with an indicator for sex, an indicator for marital status interacted with sex, a non-white indicator, a part-time indicator, and indicators for four educational attainment categories.\textsuperscript{19} The estimated $\omega_{ist}$ coefficient is our measure of the adjusted log wage in state $s$ and year $t$. Adjusted wage growth is $\Delta w_{ist} = w_{ist} - w_{ist-1}$.

Figure 4 compares our ORG-based wage growth measure to four standard measures of annual wage growth. Three of the measures, AHE, Hourly Comp, and the ECI were discussed in the previous section. The fourth is a version of the ECI that is limited to the wage and salary components of employment cost. To facilitate comparison to the other measures, the ORG-based data in figure 4 are simple means, rather than the demographically adjusted figures discussed above. The correlation of our ORG-based measure is at least 0.72 with each of the other measures. This is about as high as the other measures are correlated with each other.

Close inspection of figure 4 suggests that our ORG-based measure is most similar to the ECI wages-only measure. This is true as well in figure 5, which plots the cumulative growth in the five measures since 1979.\textsuperscript{20} The similarity of our ORG-based measure to the wages-only ECI likely reflects the fact that both measures capture only the value of wages and salaries. Neither reflects the value of benefits such as health insurance, whose relative growth rates have varied significantly over time. The AHE measure also excludes the value of benefits. Its divergence from the wages-only ECI and our ORG-based measure may be explained by its limitation to production and nonsupervisory workers.

The ORG data are our preferred source of state-level wage data. Their main attractions are large sample sizes and relatively rich associated demographic data. The lack of information on the value of benefits is a potential limitation. However, it seems plausible that the difference in growth rates
between our measure and a more inclusive measure of total compensation is constant across states in a given year. If this is the case, as we explain further below, our estimates of the sensitivity of wage growth to unemployment will be unaffected. Nevertheless, to provide a check on the sensitivity of our results to the value of benefits, we also make use of the regional detail of the ECI. Unfortunately, the ECI is reported for only four regions, which severely limits the available degrees of freedom. Moreover, we did not have access to any micro data for the ECI, so we cannot demographically adjust the data.

Finally, another limitation of the ORG data is that they are not available prior to 1979, which might be considered a relatively short time series. Thus, in order to provide some evidence on the sensitivity of wage growth to unemployment in earlier years, we also use the annual demographic files from the March CPS. These contain responses to questions on earnings, weeks worked, and usual hours per week in the previous calendar year. Thus, a wage rate can be calculated as annual earnings divided by the product of weeks worked and usual hours per week. These data are available in convenient electronic form starting in 1964, though prior to 1977, data from smaller states are not identified separately, reducing the number of degrees of freedom available. Another drawback of the March data is the smaller sample size. Nationally, the sample is around 50,000 households, but for small states, samples can be as small as a few hundred households. This tends to make the associated wage measures quite volatile from year to year. In addition, we are forced to drop some of the early years of data because of unreasonably large changes in adjusted wages that we expect are the result of changes in sample design.

**Empirical results**

Our analysis is based on a standard panel data statistical model for the response of wage growth to unemployment. That model can be written as

$$\Delta w_{st}^* = \alpha_s + \gamma_t + \beta u_t + \epsilon_{st}.$$  

where $$\Delta w_{st}^*$$ is the adjusted wage growth and $$u_t$$ is the log of the average of the 12 monthly unemployment rates for state $$s$$ in year $$t$$. The state-specific effects, $$\alpha_s$$, control for additional characteristics that are constant across time within a given state. Such factors may include demographic and industrial mix variables, as well as differences across states in the generosity of social insurance and other factors that affect the natural rate of unemployment in a given state. The year-specific effects, $$\gamma_t$$, control for the level of expected inflation in year $$t$$, as well as for the effects of productivity and other variables that may affect wages to the extent that such variables are constant across states for a given year.

Year-specific effects may also control for the effects of the exclusion of the value of benefits from our ORG-based measure of wage growth. Specifically, suppose that equation 2 holds for a comprehensive measure of compensation growth that includes the value of benefits, and further that the difference between such a measure and our ORG-based measure of wage growth is constant across states for a given year. Then $$\Delta w_{st} = \Delta w_{st}^* + g_t$$ and equation 2 can be written as

$$\Delta w_{st} = \alpha_s + \gamma_t' + u_t \beta + \epsilon_{st},$$

where $$\gamma_t' = \gamma_t + g_t$$. In this case, the lack of benefits information affects the estimates of the year effects, but not the estimate of $$\beta$$, the sensitivity of wage growth to unemployment. Moreover, if we can identify the true wage growth averaged over all states for a year with a measure such as Hourly Comp, we can adjust the estimates of the year effects to be consistent with such data. That is, $$g_t = \Delta w_{st} - \Delta \bar{w}_t$$ which is the difference between the ORG-based measure and hourly compensation for annual data.

Least-squares estimation of equation 3 is equivalent to least-squares estimation of

$$\Delta w_{st} = \tilde{\alpha}_s \beta + \epsilon_{st},$$

where $$\Delta \tilde{w}_{st} = \Delta w_{st} - \Delta \bar{w}_t - \Delta \bar{w}_s + \Delta \bar{w}$$ and $$\tilde{\alpha}_s = u_{st} - \tilde{u}_s - \tilde{u}_t + \tilde{u}$$ represent deviations from state-specific and year-specific means. That is,
We prefer the latter method of estimation for its high degree of efficiency in the face of the kind of heavy-tailed data that we employ in this article. The first two digits of the estimates of the overall sensitivity of wage growth to unemployment are unaffected by choice of estimation method. However, consistent with its greater efficiency in the presence of outliers, the estimated standard errors from the robust regression technique are slightly smaller than those for ordinary or employment-weighted least squares.

Before examining how the estimates vary over time, it is informative to look more closely at the nature of the cross-state evidence. Figure 7 shows the 1999 level of unemployment in each of the 50 states and the District of Columbia. Rates varied from a low of 2.6 percent in New Hampshire to a high of 6.5 percent in the District of Columbia. But, as we have argued previously, the simple level of unemployment in the year may not be the best guide to the tightness of a state labor market. Average unemployment rates over the 1980–99 period varied from a low of 4.0 percent in South Dakota to a high of 10.2 percent in West Virginia. Much of this variation in states’ average unemployment can be explained by slowly changing variables such as demographic composition, industry mix, and employment policies that do not necessarily affect optional wage growth.

Figure 8 shows the deviations of 1999 state unemployment rates from their averages over the 1980–99 period. These relative unemployment indicators clearly differ a good deal from the standard measures shown in figure 7. For instance, the two extremes of 1999 unemployment, New Hampshire and the District of Columbia, are reasonably similar in terms of their deviations from their average rates, being 1.8 and 1.5 percentage points lower than their averages in 1999. In terms of unemployment deviations, the tightest labor market is Michigan’s, where the 1999 unemployment rate of 3.7 percent is 4.7 points lower than its 1980–99 average of 8.4 percent. In contrast, the least tight labor market is in Hawaii where the current 5.5 percent unemployment rate is 0.4 points above its average over the last 20 years. We find that such deviations from mean unemployment rates provide a superior guide to where labor markets are tight and, thus, that the raw unemployment rates seen in figure 7 can be somewhat misleading about where wage growth should be expected to be more rapid.

\[ \Delta \bar{w}_t \] is the mean adjusted wage growth over all years in the sample for state \( s \), \( \Delta w_t \) is the mean adjusted wage growth over all states for year \( t \), and \( \Delta w \) is the overall mean of wage growth, and similarly for \( \bar{u}_s \), \( \bar{u}_t \), and \( \bar{u} \).

Figure 6 is a scatter plot of \( \Delta \bar{w}_t \) versus \( \bar{u}_t \) and thus shows the nature of the evidence on which the cross-state approach draws. A loose, but clearly negative association is apparent in the data. As shown in the first column of the first row of table 1, the ordinary least squares estimate of the regression line in figure 6 has slope \(-0.042\) with a standard error of 0.004. As in the previous scatter plots, the hyperbolic lines around the regression line represent confidence intervals for the mean wage growth associated with any level of the unemployment rate deviation. These are somewhat tighter than in the equivalent time-series scatter plots, reflecting the greatly increased degrees of freedom obtained by working with the state-level data.

Though the evidence of association seen in figure 6 is very strong, there is also a very wide scatter of points around the line. Clearly, a great many factors affect wages besides unemployment rates. Moreover, some of the very wild data points likely reflect substantial measurement error in the wage growth measure.

The second and third columns of table 1 present alternative estimation methods that reduce the influence of outliers. The second column simply weights the observations by state employment while the third column estimates the parameters using the biweight robust regression technique. We prefer the latter method of estimation for its high degree of efficiency in the face of the kind of heavy-tailed data that we employ in this article. The first two digits of the estimates of the overall sensitivity of wage growth to unemployment are unaffected by choice of estimation method. However, consistent with its greater efficiency in the presence of outliers, the estimated standard errors from the robust regression technique are slightly smaller than those for ordinary or employment-weighted least squares.

Before examining how the estimates vary over time, it is informative to look more closely at the nature of the cross-state evidence. Figure 7 shows the 1999 level of unemployment in each of the 50 states and the District of Columbia. Rates varied from a low of 2.6 percent in New Hampshire to a high of 6.5 percent in the District of Columbia. But, as we have argued previously, the simple level of unemployment in the year may not be the best guide to the tightness of a state labor market. Average unemployment rates over the 1980–99 period varied from a low of 2.6 percent in New Hampshire to a high of 6.5 percent in the District of Columbia. But, as we have argued previously, the simple level of unemployment in the year may not be the best guide to the tightness of a state labor market. Average unemployment rates over the 1980–99 period varied from a low of 4.0 percent in South Dakota to a high of 10.2 percent in West Virginia. Much of this variation in states’ average unemployment can be explained by slowly changing variables such as demographic composition, industry mix, and employment policies that do not necessarily affect optional wage growth.

Figure 8 shows the deviations of 1999 state unemployment rates from their averages over the 1980–99 period. These relative unemployment indicators clearly differ a good deal from the standard measures shown in figure 7. For instance, the two extremes of 1999 unemployment, New Hampshire and the District of Columbia, are reasonably similar in terms of their deviations from their average rates, being 1.8 and 1.5 percentage points lower than their averages in 1999. In terms of unemployment deviations, the tightest labor market is Michigan’s, where the 1999 unemployment rate of 3.7 percent is 4.7 points lower than its 1980–99 average of 8.4 percent. In contrast, the least tight labor market is in Hawaii where the current 5.5 percent unemployment rate is 0.4 points above its average over the last 20 years. We find that such deviations from mean unemployment rates provide a superior guide to where labor markets are tight and, thus, that the raw unemployment rates seen in figure 7 can be somewhat misleading about where wage growth should be expected to be more rapid.
Table 1 also shows estimates of the response of wage growth to unemployment for four five-year periods. The results suggest that wage growth has become somewhat less sensitive to unemployment in the 1990s. The robust regression methodology yields estimates of –0.045 and –0.044 for the early and late 1980s. The coefficient estimate for the early 1990s fell to –0.039, and that for the late 1990s was –0.033. Of course, even in the late 1990s, the estimates in table 1 are highly statistically significant, with t-statistics of around five. There is modestly strong evidence that the coefficient has changed over time. The F statistics shown in the table imply that the hypotheses that the 1995–99 coefficient is the same as the 1980–84, 1985–89, and the 1980–94 averages can be rejected at the 10 percent level, but not at the 5 percent level. The hypothesis that the 1995–99 coefficient is the same as the 1990–94 coefficient cannot be rejected at any standard confidence level.

Figure 9 shows the result of estimating a separate slope for each year of the sample. Such estimates are based on the model

\[ \Delta w_{it} = \alpha + \gamma + \beta + \epsilon_{it}, \]

which continues to impose a common state effect, but allows the intercept and slope to vary freely over the sample period. Robust estimates of the slopes by year are plotted in figure 9 along with 90 percent confidence intervals. Since each data point is essentially estimated from 51 rather noisy observations, the confidence intervals tend to be somewhat wide. Still, all 20 coefficients are statistically significant at the 5 percent level.

The pattern of estimates shown in figure 9 leads us to view the evidence of a systematic drop in the magnitude of the coefficient as somewhat weak. The magnitude of the elasticity has decreased in recent years, with 1998 having the single smallest coefficient. But as recently as 1994 and 1995 the coefficient was about as large as it ever has been. And there have been previous years—1985 and 1993—in which the coefficient has declined, only to increase again subsequently.

The drop in coefficients in table 1 is also dependent on the imposition of a constant elasticity functional form. Such

**TABLE 1**

<table>
<thead>
<tr>
<th></th>
<th>OLS</th>
<th>WLS</th>
<th>Robust</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log unemployment rate</td>
<td>–0.042*</td>
<td>–0.042*</td>
<td>–0.042*</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.004)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>Adjusted R-squared</td>
<td>0.467</td>
<td>0.550</td>
<td>0.463</td>
</tr>
<tr>
<td>Unemployment rate, 1980–84</td>
<td>–0.047*</td>
<td>–0.049*</td>
<td>–0.045*</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.005)</td>
<td>(0.005)</td>
</tr>
<tr>
<td>Unemployment rate, 1985–89</td>
<td>–0.046*</td>
<td>–0.046*</td>
<td>–0.044*</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.005)</td>
<td>(0.005)</td>
</tr>
<tr>
<td>Unemployment rate, 1990–94</td>
<td>–0.038*</td>
<td>–0.040*</td>
<td>–0.039*</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.007)</td>
<td>(0.006)</td>
</tr>
<tr>
<td>Unemployment rate, 1995–99</td>
<td>–0.032*</td>
<td>–0.030*</td>
<td>–0.033*</td>
</tr>
<tr>
<td></td>
<td>(0.007)</td>
<td>(0.006)</td>
<td>(0.006)</td>
</tr>
</tbody>
</table>

F test p-statistic:
- UR, 1980–94=UR, 1995–99: 0.082, 0.027, 0.086
- UR, 1980–84=UR, 1995–99: 0.059, 0.011, 0.074
- UR, 1985–89=UR, 1995–99: 0.058, 0.037, 0.092
- UR, 1990–94=UR, 1995–99: 0.435, 0.218, 0.395

Adjusted R-squared: 0.469, 0.552, 0.461

*significant at the 5 percent level.

Notes: OLS is the ordinary least squares estimate. WLS is the observation weighted by state employment. UR is the unemployment rate. All regressions include state and year fixed effects. The last column includes industry and occupational composition controls. Robust standard errors are in parentheses. The unemployment rate for each period is measured by the log of the unemployment rate times a dummy variable for the time period. The F test measures are calculated by the log of the unemployment rates times the dummy variable for one period being held equal to the log of the unemployment rate times the dummy variable for another period. ORG wage data are adjusted for education, experience, gender, race, and full time status.

Sources: Authors’ calculations using data from the U.S. Department of Labor, Bureau of Labor Statistics for the unemployment rate and from the U.S. Department of the Census, Current Population Survey, for the weighted averages from the ORG for industry, occupation, and union composition.
Rather, the late 1980s appears to be the period that was different, having a higher estimated coefficient than the other three periods. We prefer the constant elasticity specification of table 1 because of the better fit to the data, but the results of table 2 reinforce our view that the evidence of a decline in the sensitivity of wage growth to unemployment is rather weak.

Table 3 explores the sensitivity of the results in table 1 to alternative specifications. These all employ the robust regression methodology, but change other aspects of the specification. The first column shows the slope coefficients when we include additional variables measuring the fraction of workers in the various one-digit industries and occupations. Such variables may control for variation across states in productivity growth and other factors that determine wage growth. The coefficients tend to be smaller in magnitude than those in table 1, but the conclusions one would draw are similar, while the coefficient for the late 1990s is somewhat smaller, it is still highly statistically significant.

**TABLE 2**

State wage curve elasticities, alternative labor market indicators

<table>
<thead>
<tr>
<th>Log wage on level of labor market condition</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Unemployment rate, 1980–84</td>
<td>-0.0053*</td>
</tr>
<tr>
<td></td>
<td>(0.0007)</td>
</tr>
<tr>
<td>Unemployment rate, 1985–89</td>
<td>-0.0068*</td>
</tr>
<tr>
<td></td>
<td>(0.0007)</td>
</tr>
<tr>
<td>Unemployment rate, 1990–94</td>
<td>-0.0063*</td>
</tr>
<tr>
<td></td>
<td>(0.0010)</td>
</tr>
<tr>
<td>Unemployment rate, 1995–99</td>
<td>-0.0064*</td>
</tr>
<tr>
<td></td>
<td>(0.0012)</td>
</tr>
</tbody>
</table>

F test p-statistic:

- UR, 1980–94=UR, 1995–99 0.751
- UR, 1980–84=UR, 1995–99 0.358
- UR, 1985–89=UR, 1995–99 0.760
- UR, 1990–94=UR, 1995–99 0.988
- Adjusted R-squared 0.450

*Significant at the 5 percent level.

Notes: UR is the unemployment rate. Regression includes state and year fixed effects and is estimated using robust regression. The unemployment rate for each period is measured by the log of the unemployment rate times a dummy variable for the time period. The F test measures are calculated by the log of the unemployment rates times the dummy variable for one period being held equal to the log of the unemployment rate times the dummy variable for another period.
The next column in table 3 uses the unemployment rate from the year before rather than the current year. This lowers the coefficients. The decline in the recent period is smaller, however. The next three columns explore the sensitivity of the results to the inclusion of fixed effects. Leaving out year effects makes the coefficients larger in magnitude, reflecting the fact that years with lower unemployment have had higher than average wage growth. Leaving out state effects significantly weakens the results, which reflects the fact that states with higher than average mean unemployment rates tend to have higher mean wage growth. Leaving out both kinds of fixed effects produces weak results as well. Both kinds of fixed effects are statistically significant according to the usual F statistic. Thus we prefer the specification estimated in table 1, and view the other results as indicating the effects of various forms of specification errors. Finally, fitting a nonlinear specification seems to us to be asking a lot of 51 noisy observations. Clearly, figure 6 shows that there is a wide scatter around what is still a highly significant negative relationship. Thus, it would be quite surprising to see a clean pattern of means across quartiles when each of those means was estimated with only 12 or 13 observations.

As we have noted, Lehrman and Schmidt (1999) report no evidence of a cross-state association between unemployment and wage growth. The results above may explain some of the difference between their results and ours. Lehrman and Schmidt use the unemployment rate for only the last quarter of the period, rather than the average over the whole period. The results in table 3 using lagged unemployment rates suggest that the match of the time periods of unemployment and wage growth may matter. Lehrman and Schmidt also use data on unemployment in 1998, which figure 9 says provides the weakest results of any year. Moreover, they only look at a single cross-section of data and so cannot control for state-specific fixed effects which table 3 shows is important. Finally, fitting a nonlinear specification seems to us to be asking a lot of 51 noisy observations. Clearly, figure 6 shows that there is a wide scatter around what is still a highly significant negative relationship. Thus, it would be quite surprising to see a clean pattern of means across quartiles when each of those means was estimated with only 12 or 13 observations.

One possible explanation for the falling coefficient on unemployment in table 1 is the changing nature of the work force. For instance, it is has been previously shown that wage growth among college-educated workers is less sensitive to unemployment
than that among other workers. Thus the increasing share of college-educated workers could cause a decline in the unemployment coefficient of the kind seen in table 1. The results in table 4, however, show that this is not the case. The decline in coefficients is seen both for noncollege and college workers. Something other than a compositional shift towards college workers explains the lower late-1990s coefficients on unemployment.

Table 5 shows estimates of our basic specification using the March CPS data. As we noted, the advantage of this dataset is that it is available for earlier periods. Its disadvantage is that its wage measures are noisier, being based on a sample one-third as large as the ORG data. The results shown for five year intervals between 1964 and 1998, the last available data, suggest a quite stable relationship between unemployment and wage growth, with elasticity estimates generally near –0.03 except for the 1984 to 1988 period when the elasticity was estimated to be –0.045. Moreover, the F-statistics indicate that even the latter estimate is not statistically different from the estimate for the most recent period. The coefficients in table 5 are, however, somewhat lower than those in table 1. This must reflect differences in the nature of the March CPS wage measure, which is based on the previous calendar year, rather than the previous week.

Table 6 reports results obtained from the regional ECI data both for wages and salaries only and for total compensation. Because these data are available for only four regions, there are many fewer degrees of freedom. The first and third columns show results for periods similar to those shown in table 1. These results for wages and salaries are relatively similar to those in table 1, except in the first period, when the data may have been somewhat suspect due to the newness of the series. However, for total compensation, the coefficient for the most recent five-year period is small and not statistically significant. Looking closely

| TABLE 4 |
| State wage curve elasticities, by education |
| Noncollege sample | College sample |
| Unemployment rate, 1980–84 | –0.047* (0.006) | –0.038* (0.011) |
| Unemployment rate, 1985–89 | –0.046* (0.005) | –0.037* (0.009) |
| Unemployment rate, 1990–94 | –0.039* (0.006) | –0.034* (0.011) |
| Unemployment rate, 1995–99 | –0.035* (0.006) | –0.027* (0.011) |
| F test p-statistic: | | |
| UR, 1980–94=UR, 1995–99 | 0.111 | 0.371 |
| UR, 1980–84=UR, 1995–99 | 0.083 | 0.395 |
| UR, 1985–89=UR, 1995–99 | 0.087 | 0.397 |
| UR, 1990–94=UR, 1995–99 | 0.531 | 0.594 |
| Adjusted R-squared | 0.424 | 0.202 |

*Significant at the 5 percent level.

Notes: UR is the unemployment rate. All regressions include state and year fixed effects and, unless noted, are estimated using robust regression. The unemployment rate for each period is measured by the log of the unemployment rate times a dummy variable for the period. The F test measures are calculated by the log of the unemployment rates times the dummy variable for one period being held equal to the log of the unemployment rate times the dummy variable for another period.

| TABLE 5 |
| State wage curve elasticities |
| Wage growth: March CPS, 1964–98 |
| Unemployment rate, 1964–68 | –0.028* (0.013) |
| Unemployment rate, 1969–73 | –0.026 (0.014) |
| Unemployment rate, 1974–78 | –0.033* (0.012) |
| Unemployment rate, 1979–83 | –0.030* (0.009) |
| Unemployment rate, 1984–88 | –0.045* (0.007) |
| Unemployment rate, 1989–93 | –0.028* (0.009) |
| Unemployment rate, 1994–98 | –0.030* (0.009) |
| F test p-statistic: | | |
| UR, 1964–93=UR, 1994–98 | 0.656 |
| UR, 1964–68=UR, 1994–98 | 0.906 |
| UR, 1969–73=UR, 1994–98 | 0.813 |
| UR, 1974–78=UR, 1994–98 | 0.845 |
| UR, 1979–83=UR, 1994–98 | 0.977 |
| UR, 1984–88=UR, 1994–98 | 0.164 |
| UR, 1989–93=UR, 1994–98 | 0.891 |
| Time period | 1964–98 |
| Adjusted R-squared | 0.614 |

*Significant at the 5 percent level.

Notes: UR is the unemployment rate. All regressions include state and year fixed effects and are estimated using robust regression. The unemployment rate is from the BLS for 1978–98 and state UI records for 1964–77. Some states are not uniquely identified in the March CPS prior to 1977. The unemployment rate for each period is measured by the log of the unemployment rate times a dummy variable for the time period. The F test measures are calculated by the log of the unemployment rates times the dummy variable for one period being held equal to the log of the unemployment rate times the dummy variable for another period.
at the individual observations suggests, however, that a very small number of data points are driving this result. Moreover, when we break the data into three-year intervals, the results suggest less evidence of a drop in the sensitivity of total compensation growth to unemployment. Given how little regional variation underlies the data in table 6, we consider the consistency of the results with those in table 1 to be reasonably good.

Thus far, our results have been limited to showing how the sensitivity of wage growth to unemployment has varied over time. Table 7 shows, in addition, how the level of wage growth associated with any level of unemployment has varied over time. Such quantities depend on both the estimated slope coefficients, \( \beta_t \), and the year effects, \( \gamma_t \). The values shown in table 7 are based on the specification of table 1 in which slopes are constant for each five-year period. The values in the column labeled Average Intercept–Raw are the average of the five-year effects \( \gamma \) estimated for the period. The adjusted values in the next column are our estimates of the \( \gamma' \), the values that would correspond to the more comprehensive Hourly Comp wage growth measure. The intercept values are somewhat difficult to interpret because they potentially capture the effects of a number of variables. However, the fact they have fallen over time is consistent with the notion that they capture changes in expected inflation.

Given the normalization that \( \sum \alpha_t = 0 \), the predicted mean ORG-based adjusted wage growth associated with log unemployment rate \( \bar{u} \) for year \( t \) is \( \Delta \bar{w} = \gamma' + \bar{u} \beta' \), and the predicted mean Hourly Comp growth is \( \Delta w = \gamma' + \bar{u} \beta' \). To obtain estimates of predicted real wage growth, we subtract the rate of price inflation. In particular, the estimated amount by which the growth of Hourly Comp exceeds the growth in business sector prices, which is a reasonable measure of real wage growth, is \( \Delta w = \gamma' + \Delta \bar{p} + \bar{u} \beta' \), where \( \Delta \bar{p} \) is the change in the log average price deflator for the business sector. Table 7

<table>
<thead>
<tr>
<th>State wage curve elasticity</th>
<th>Wage growth: Employment Cost Index, 1983–99</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Wages and salaries</td>
</tr>
<tr>
<td></td>
<td>5-year intervals</td>
</tr>
<tr>
<td>Unemployment rate, 1983–84</td>
<td>-0.007 (0.005)</td>
</tr>
<tr>
<td>Unemployment rate, 1985–89</td>
<td>-0.030* (0.005)</td>
</tr>
<tr>
<td>Unemployment rate, 1990–94</td>
<td>-0.039* (0.010)</td>
</tr>
<tr>
<td>Unemployment rate, 1995–99</td>
<td>-0.025* (0.007)</td>
</tr>
<tr>
<td>Unemployment rate, 1983–85</td>
<td>-0.019* (0.005)</td>
</tr>
<tr>
<td>Unemployment rate, 1986–88</td>
<td>-0.031* (0.005)</td>
</tr>
<tr>
<td>Unemployment rate, 1989–91</td>
<td>-0.035* (0.011)</td>
</tr>
<tr>
<td>Unemployment rate, 1992–94</td>
<td>-0.021* (0.010)</td>
</tr>
<tr>
<td>Unemployment rate, 1995–97</td>
<td>-0.015** (0.009)</td>
</tr>
<tr>
<td>Unemployment rate, 1998–99</td>
<td>-0.044* (0.009)</td>
</tr>
<tr>
<td>Adjusted R-squared</td>
<td>0.757</td>
</tr>
</tbody>
</table>

*Significant at the 5 percent level.
**Significant at the 10 percent level.

Notes: UR is the unemployment rate. All regressions include state and year fixed effects and are estimated using robust regression. The Employment Cost Index (ECI) is aggregated to four regions—East, South, Midwest, and West. Therefore, the sample includes four regions over 17 years, or 68 observations. ECI data are not demographically adjusted. The unemployment rate for each period is measured by the log of the unemployment rate times a dummy variable for the time period.
Federal Reserve Bank of Chicago

shows the predicted average real wage growth calculated in this manner for unemployment rates of 4 percent, 6 percent, and 8 percent. For an unemployment rate of 4 percent, predicted real wage growth dropped between the early and late 1980s, but has been reasonably constant since then. Our estimates currently predict real wage growth of 2.8 percent when the unemployment rate is 4 percent, about its current value. The predicted real wage growth rates associated with 6 percent and 8 percent unemployment also fell between the early and late 1980s, and since then have been fairly constant. The 0.6 percent level of wage growth predicted for 8 percent unemployment in the last period has, however, returned to about its level for the early 1980s.

One can also ask what level of unemployment is predicted to deliver a particular rate of real wage growth, say \( \Delta \left( \frac{w^*}{p} \right) \). According to the above, that unemployment rate is \( u^* = \left[ \Delta \left( \frac{w^*}{p} \right) - (\gamma_t - \Delta p_t) \right] / \beta_t \). The last column of table 7 shows the values of this quantity corresponding to the mean real wage growth rate over the 1980–99 period, which was about 1.5 percent per year. That unemployment rate was nearly 7 percent in the early 1980s, but has been relatively constant since then at about the 6 percent level that we estimate for the late 1990s. We view the results in table 7 as confirming the relatively stable relationship between wage growth in excess of inflation and unemployment.

We argued previously that there might be labor market variables that predict wage growth better than the standard civilian unemployment rate. The recent drop in the coefficient on unemployment seen in table 1 might even reflect a misspecification in which unemployment is proxying for a more appropriate measure of labor market conditions. The drop in the unemployment coefficient might then be due to a lower correlation of unemployment with the preferred variable, which could have a stable relationship to wage growth. The results in table 8 suggest, however, that the decline in the coefficients in table 1 are not due to the unemployment rate becoming a poorer proxy for a superior measure of labor market tightness. The table shows the results of replacing the unemployment rate with several other measures of labor market conditions. These include an unemployment rate calculated from the ORG data, a measure of unemployment that includes all nonemployed workers who say they want a job regardless of whether they have recently searched, an even broader unemployment rate that also includes those who work part-time for economic reasons, a narrower measure that includes only white males between the ages of 25 and 54, the employment-to-population ratio, a measure of the exit rate out of unemployment, the fraction of the labor force unemployed five or fewer weeks, and the portion of the labor force unemployed 15 or more weeks. Virtually all the measures show the decline in coefficient magnitude in the most recent period that we see in table 1 for the unemployment rate. The drop off in the sensitivity of wage growth is especially significant for the exit rate out of unemployment and the rate of short-term unemployment. This may reflect the introduction of computer-aided interviewing technology with the 1994 CPS redesign, which had the effect of introducing a break in the series on short-term unemployment.

The results in table 8 suggest that the standard unemployment rate is not the only measure that might be used to judge the tightness of labor market conditions. Judging by the standard R-squared measure, several variables predict wage growth about as well as the unemployment rate. Indeed, the rate of long-term unemployment actually does very slightly
Economic Perspectives

better. The two broader measures of unemployment, which include all of those who say they want a job and those workers plus those who are involuntarily part-time, come reasonably close to matching the predictive power of the standard unemployment rate, while the narrower measure that is limited to prime-age white males does less well. Perhaps somewhat surprisingly, the measures that may be more closely connected to theory, the employment-to-population ratio and the exit rate from unemployment, are among the least well performing measures, though in the latter case this may be due to breaks in the data series that may, with some work, be repairable. A fully satisfactory comparison of the forecasting abilities of the various labor market variables would require the use of higher frequency data, more elaborate dynamics, and some attention to the out-of-sample properties of the forecasts. We regard the results in table 8 as suggesting that such work may be quite fruitful.

**Conclusion**

In this article, we have shown that the negative cross-state correlation between unemployment and wage growth persists even in recent data. We find some evidence of a decline in the sensitivity of wage growth to unemployment in the late 1990s. But, we regard that evidence as being somewhat weak because it is dependent on exactly when the line between periods is drawn and whether the relationship is modeled as one in which percentage or absolute differences in unemployment rates have constant effects on wage growth.

Of course, the relationship between unemployment and wage growth is a loose one. Unemployment is only one of many factors that affect wage growth, so that looking at a small number of states or years, differences in unemployment rates may not always provide a good prediction of differences in wage growth. But with enough data, the relationship between

---

**TABLE 8**

State wage curve elasticities, alternative labor market indicators

<table>
<thead>
<tr>
<th></th>
<th>BLS unemployment rate</th>
<th>ORG unemployment rate</th>
<th>Unempl plus NILF who want job</th>
<th>Unempl plus NILF who want job plus PT for econ reasons</th>
<th>White male age 25–54 unemployment rate</th>
<th>Empl-pop ratio*</th>
<th>Exit rate out of unemployment</th>
<th>Unempl 0–5 weeks*</th>
<th>Unempl 15+ weeks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unemployment rate,</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1980–84</td>
<td>-0.045*</td>
<td>-0.043*</td>
<td>-0.050*</td>
<td>-0.058*</td>
<td>-0.024*</td>
<td>0.194*</td>
<td>0.036*</td>
<td>0.025*</td>
<td>-0.022*</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.005)</td>
<td>(0.006)</td>
<td>(0.007)</td>
<td>(0.004)</td>
<td>(0.029)</td>
<td>(0.007)</td>
<td>(0.008)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>1985–89</td>
<td>-0.044*</td>
<td>-0.042*</td>
<td>-0.047*</td>
<td>-0.051*</td>
<td>-0.023*</td>
<td>0.173*</td>
<td>0.025*</td>
<td>0.026*</td>
<td>-0.022*</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.005)</td>
<td>(0.005)</td>
<td>(0.005)</td>
<td>(0.003)</td>
<td>(0.028)</td>
<td>(0.007)</td>
<td>(0.007)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>1990–94</td>
<td>-0.039*</td>
<td>-0.035*</td>
<td>-0.039*</td>
<td>-0.038*</td>
<td>-0.021*</td>
<td>0.164*</td>
<td>0.022*</td>
<td>0.001</td>
<td>-0.020*</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.006)</td>
<td>(0.007)</td>
<td>(0.007)</td>
<td>(0.004)</td>
<td>(0.030)</td>
<td>(0.006)</td>
<td>(0.008)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>1995–99</td>
<td>0.033*</td>
<td>-0.027*</td>
<td>-0.031*</td>
<td>-0.029*</td>
<td>-0.016*</td>
<td>0.176*</td>
<td>0.001</td>
<td>0.003</td>
<td>-0.014*</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.006)</td>
<td>(0.006)</td>
<td>(0.007)</td>
<td>(0.004)</td>
<td>(0.030)</td>
<td>(0.002)</td>
<td>(0.006)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>F test p-statistic:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UR, 1980–94=UR,</td>
<td>0.086</td>
<td>0.023</td>
<td>0.024</td>
<td>0.002</td>
<td>0.055</td>
<td>0.936</td>
<td>0.000</td>
<td>0.001</td>
<td>0.020</td>
</tr>
<tr>
<td>1995–99</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UR, 1980–94=UR,</td>
<td>0.074</td>
<td>0.022</td>
<td>0.015</td>
<td>0.001</td>
<td>0.058</td>
<td>0.436</td>
<td>0.000</td>
<td>0.002</td>
<td>0.043</td>
</tr>
<tr>
<td>1995–99</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UR, 1985–89=UR,</td>
<td>0.092</td>
<td>0.024</td>
<td>0.026</td>
<td>0.003</td>
<td>0.094</td>
<td>0.897</td>
<td>0.002</td>
<td>0.000</td>
<td>0.034</td>
</tr>
<tr>
<td>1995–99</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UR, 1990–94=UR,</td>
<td>0.395</td>
<td>0.252</td>
<td>0.337</td>
<td>0.274</td>
<td>0.246</td>
<td>0.607</td>
<td>0.001</td>
<td>0.637</td>
<td>0.138</td>
</tr>
<tr>
<td>1995–99</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adjusted R-squared</td>
<td>0.461</td>
<td>0.453</td>
<td>0.448</td>
<td>0.457</td>
<td>0.438</td>
<td>0.409</td>
<td>0.413</td>
<td>0.412</td>
<td>0.466</td>
</tr>
</tbody>
</table>

*Significant at the 5 percent level.

*Detrended.

1994 is excluded.

Notes: UR is the unemployment rate. ORG is the outgoing rotation groups. BLS indicates U.S. Bureau of Labor Statistics. NILF is not in labor force. PT indicates part time. All regressions include state and year fixed effects and are estimated using robust regression. The unemployment rate for each period is measured by the log of the unemployment rate times a dummy variable for the time period. The F test measures are calculated by the log of the unemployment rates times the dummy variable for one period being held equal to the log of the unemployment rate times the dummy variable for another period.
unemployment and wage growth emerges fairly clearly and does not appear to be dependent on any arbitrary details of our analysis.

We also find that several other labor market indicators predict wage growth about as well as the standard civilian unemployment rate. Refining such measures and studying their forecasting abilities more systematically may be a fruitful area for further research.

Finally, our results may have implications for work on inflation forecasting, an important component in the monetary policy process. Traditional short-run, or expectations-augmented, Phillips curve methodologies have tended to overpredict the change in inflation in recent years. That methodology depends upon both the relationship between unemployment and expected wage growth and the relationship between wage growth and price inflation. Given the many fundamental changes that may be affecting the labor market, it is natural to look for a change in the relationship between unemployment and wage growth. But, our finding that the cross-state relationship between unemployment and wage growth has been relatively stable suggests that more attention be given to the link between wage growth and price inflation as the source of instability in the short-run Phillips curve. This seems consistent with findings such as those in Brayton et al. (1999) that adding variables to account for variation in the markup of prices over wages may be the most attractive way to stabilize the relationship between unemployment and changes in price inflation.

NOTES

1Friedman (1968) and Phelps (1973) are classic statements of this point.

2In the years since Phillips’ (1958) paper, the correlation between nominal wage growth and unemployment has been close to zero in U.S. data.

3Abraham et al. (1999) discuss the differences in these wage measures.

4Blanchard and Katz (1997) discuss the relationship between the kind of time-series evidence depicted in figures 1–3 and the cross-state evidence that is the main focus of this article.

5Blanchard and Katz (1997) note that, empirically, these other variables are often found to have little impact on wage growth forecasts.

6These were computed under the usual ideal assumptions that error terms are uncorrelated and of constant variance, and thus may be somewhat optimistic. The hyperbolic lines around the regression line represent 90 percent confidence intervals for the expected level of wage growth in excess of inflation at a given level of log unemployment.


8See, for example, Mortensen and Pissarides (1994).

9See, for example, Shapio and Stiglitz (1984) and Salop (1979).

10Blanchard and Katz (1997) provide a cogent discussion of these issues.

11Castillo (1998) shows that in U.S. data, those outside the labor force who want a job are less attached to the labor market than unemployed workers. However, Jones and Riddel (1999) show that in Canadian data, those out of the labor force who report wanting a job are closer to the unemployed than to others who are out of the labor force, in terms of their subsequent probabilities of employment.

12An important reference is Blanchflower and Oswald (1994), who document a cross-sectional relationship between unemployment and wages in a number of countries over a number of periods. Blanchflower and Oswald interpret their results as a relationship between unemployment and the level of wages because in their statistical models for the wage level, lagged wages are estimated to have small coefficients. We agree, however, with Blanchard and Katz (1997) and Card and Hyslop (1996) that these low estimates are the result of substantial measurement error in Blanchflower and Oswald’s wage measures as well as their inappropriate use of annual, rather than hourly earnings. We find that in models employing hourly wage measures obtained from samples large enough to minimize measurement error, the coefficient on lagged wages is quite close to unity. Thus, the relationship is best thought of in terms of wage growth rather than wage levels. Roberts (1999) and Whelan (1999) show that the form of the micro-data relationship may not matter for the form of aggregate inflation dynamics.

13Results on wage growth across states are a small part of Lehman and Schmidt’s (1999) lengthy study. The description of the empirical analysis in Zandi (2000) is not particularly detailed, but his results appear to be consistent with our findings. Zandi concludes that the Phillips curve is “alive and kicking.” Whether this follows from his or our evidence depends, however, on what one means by the “Phillips curve.” If one means that expected wage growth is related to unemployment, we agree with his conclusion. However, as we discuss below, if the Phillips curve is taken to be the short-run, or expectations-augmented, relationship between unemployment and changes in price inflation, his conclusion doesn’t necessarily follow from his results.

14See, for example, Gordon (1997).

15See, for example, Blanchard and Katz (1997).

16See Brayton et al. (1999).

17Until 1996, there were approximately 60,000 households in the survey.
We drop observations on workers whose computed wage is less than 50 cents per hour or more than $100 per hour.

Blanchard and Katz (1997) estimate separate regression models for each year of data while we estimate a single, pooled regression. This makes no appreciable difference to the results when, as in the models we estimate, year effects are included in the estimation.

The ECI compensation series is scaled to equal the ORG measure in 1982, the first year it is available.

Prior to 1976, data on usual weekly hours is not available; in its place we use data on hours worked in the week prior to the survey.

In our analysis of the March data, unemployment rates before 1978 are obtained from state unemployment insurance claims data.

This argument goes through more generally if the difference between the ORG wage growth measure and an ideal wage growth measure has an error components structure that is limited to a year effect, a state effect, and an error term that is uncorrelated with unemployment.

We use the default tuning parameters in the Stata statistical procedure. (These control the rate at which outliers are down-weighted.) See Stata Corporation (1999) for a description of the technique.

In a cross-sectional regression, such variables explain about 60 percent to 75 percent of the variation in state average unemployment rates.

Hawaii is the only state for which 1999 was an above-average year for unemployment.

The regional ECI data are not available before 1983.

See, for example, Brayton et al. (1999).

REFERENCES


