Wages, productivity, employment, inflation and distribution are linked by accounting identities.

Labor productivity (or just productivity, for short) is defined as the amount produced, divided by the amount of labor used to produce it. The labor share is defined as the fraction of total income paid out to labor. We can use these two **identities** to analyze changes in output, employment, wages and prices. This doesn't tell us what will happen in the economy, but it does tell us something about what *can* happen. Using these identities also helps us describe developments in the economy more precisely, and clarifies what assumptions are needed for various stories or predictions about the economy to be true.

Accounting identity. An equation that must always be true, because of how the terms are defined.

There is a mathematical rule that lets us convert equations to linear form, which is easier to work with.

A **linear equation** is one in which the variables are only added or subtracted. None of the variables are multiplied by each other, and none are raised to a power (that is, there are no expressions like x^2). Linear equations are generally easier to work with, so it's convenient to be able to change other kinds of equations to linear ones if possible.

Linear equation. An equation in which the terms are only added or subtracted. None of the variables are multiplied or divided, and none have exponents.

One useful tool for making linear equations is: If $a = b * c$ then

$$\textit{percentage change in a} \approx \textit{percentage change in b} + \textit{percentage change in c} \tag{1}$$

Similarly, if $a = b/c$ then

$$\textit{percentage change in a} \approx \textit{percentage change in b} - \textit{percentage change in c} \tag{2}$$

The Greek letter Δ (delta) is often used to mean the change in a variable. So to save space, I will write $\%\Delta$ when I mean "percent change in..." For example " $\%\Delta$ employment" means "percent change in employment."

So we can rewrite Equation 1 as:

$$\%\Delta a \approx \%\Delta b + \%\Delta c$$

This is linear – the variables are simply added. Whereas $a = b * c$ is not linear, since the variables are multiplied. Note that the original equation described the *levels* of the variables, while the new, linear one describes the *changes* in them.

This works the same if we have more than two variables on the right hand side.

Labor productivity is defined as output divided by the amount of labor employed.

When economists talk about "productivity", they mean either **labor productivity** or **total factor productivity**. Labor productivity is the output produced by a given amount of labor; total factor productivity is the amount of output produced by a given amount of labor and capital. Total factor productivity is important for economic theory, but it is hard to apply in practice, since measuring capital is difficult and you need to make additional assumptions about how the labor and capital are combined. For most practical purposes, labor productivity is more relevant. Whenever someone refers to "productivity" by itself, they almost always mean labor productivity.

Labor productivity. Total output divided by total employment. The most common measure of productivity.

Total factor productivity. Total output divided by the labor and capital used. Important in economic theory but hard to apply in the real world.

Labor productivity is defined as output divided by the amount of labor used. Here we will measure labor by number of people employed. So labor productivity is defined by:

$$productivity = \frac{output}{employment} \tag{3}$$

We can measure productivity for the economy as a whole, for an industry or sector, or for a single business. If we are measuring it for the economy as a whole, then "output" is GDP; for an industry or business, it is value added. When we are talking about changes in productivity, we normally measure output in real (or inflation-adjusted) terms.

We can rearrange Equation 3 to get

$$output = employment * productivity$$

In other words, total production in an economy (or an industry or business) is equal to the number of people employed, times the average amount produced by each one.

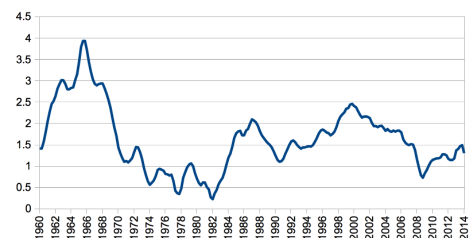


Figure 1: US labor productivity growth rates, 5-year moving averages. In recent years, productivity has grown at less than 1.5 percent per year, which is lower than in much of the postwar period.

Labor can be measured as either the number of people employed, or the number of hours of work.

Labor can be measured either in hours of work, or number of people employed. For the modern US, which we use will not have much effect on measured productivity growth, since the average number of hours worked over a week or a year have changed very little in recent decades. But if we want to compare productivity between countries, how we measure labor can make a big difference. For example, in the US, the average employed person works 1,800 hours a year, a number that has not changed since the 1970s. In France over the same period average working hours per year have fallen from 1,800 to 1,500. As of 2017, if you measure productivity as output per worker, productivity in the US appears to be about 20 percent higher than in France. But if you measure it as output per hour, productivity in the US and France are almost exactly equal.[1]

[1] Thomas Piketty, "Of productivity in France and in Germany, January 2017.

The change in employment over some period of time is equal to the change in output minus *the change in labor productivity.*

We can analyze changes in employment in terms of changes in output and productivity. Using Equation 2, we can write:

$$\%\Delta employment \approx \%\Delta output - \%\Delta productivity \qquad (4)$$

The percent change in employment is equal to the percent change in output minus the percent change in employment. For example, in 2014, total employment in the US rose by 2.2 percent, output rose by 2.9 percent, and productivity rose by 0.7 percent. (This is an exceptionally low rate by historical standards.) We can apply this equation to a change over one year or over several years. But if we apply it to a very long period (say, 50 years), the approximation may be less accurate.

Equation 4 is an accounting identity: it is true by definition. But it still shows us a couple of things that might not be obvious.

First of all, changes in employment can be due to either changes in total output, or to changes in productivity – that is, either changes in how much is produced, or in how much labor is used for a given amount of production. Over short periods, changes in output growth are much more important. For example, Table 1 shows the average annual change in employment during the expansion of 2002-2007 and the recession of 2008-2009.

Period	Employment	Output	Productivity
2002-2007	0.8%	2.7%	1.9 %
2008-2009	-3.1%	-1.5%	1.6%

Table 1: Average Annual Change in Employment, Output and Productivity

Employment grew at an average rate of 0.8 percent per year over 2002-2007, and fell at a rate of 3.1 percent per year during 2008-2009. This difference is entirely explained by the fact that output was rising during the first period, and falling in the second period. As you can see, labor productivity actually grew somewhat slower during the recession than during the expansion, but the change is quite small compared with the changes in output and employment growth.

Over long periods, faster labor productivity growth could contribute to slower employment growth. This is called "technological unemployment," but it is not clear that it is a real problem.

Over longer periods, however, changes in the speed of labor productivity growth may be more important. The second thing that Equation 4 tells us is that if output is growing at a constant rate, than faster productivity growth must men slower growth in employment. If productivity grew fast enough, you might even see a situation where output continued to grow while employment fell.

The idea that rapid improvements in labor productivity might lead to a fall in employment is a familiar one in the media and in policy discussions. It is often referred to as **technological unemployment** – the idea that "robots will take our jobs." Obviously, there are many specific cases of work jobs that become obsolete through technological change. But Equation 4 helps us think about this possibility more systematically. To begin with, it highlights the point that a prediction that technological progress will reduce employment is simply a claim that we will see an acceleration of productivity growth but not of GDP growth. But this raises two questions. First, productivity growth has been slowing down in recent years, not accelerating. Since 2000, productivity growth has averaged barely one percent a year, compared with around 1.5 percent per year during the period between 1950 and 2000 (and as high as 3 percent a year during the 1960s). So the "robots will take our jobs" story is not just extrapolating from what is already happening; someone telling this story has to explain why the recent trend of declining productivity will reverse itself. Second, Equation 4 makes it clear that faster productivity growth can lead to lower employment *or* to faster growth in output. Someone telling the "robots will take our jobs" story also has to explain why faster productivity growth will not simply lead to faster growth of GDP. After all, 120 years ago most Americans worked in agriculture. Technological change has resulted in the disappearance of almost all of those jobs. But the result has not been mass unemployment, but rather increased production in the other (secondary and tertiary) sectors of the economy.

Technological unemployment. Unemployment that results from labor productivity rising faster than total output, so that fewer workers are needed.

While the statistics are not decisive either way, there is some evidence that labor productivity rises faster when output is growing more rapidly. In this case, we might see the opposite of technological unemployment – employment and productivity moving together. For example, the early 1930s, when employment fell very steeply, labor productivity actually declined – one of the only periods on record when this occurred. And when employment rose in the recovery from the Depression, productivity rose as well. Note that in this case

Equation 4 was still true – as an accounting identity, it is always true – but the big changes in output overwhelmed the effect of productivity on employment.

The technological unemployment issue is an example of why accounting identities are useful. They can't prove that a certain story about the economy is true. But they can clarify what that story means, and show what assumptions it involves.

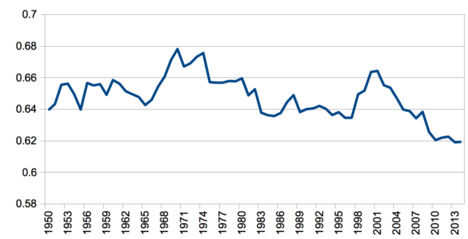


Figure 2: US labor share. During the 1980s, and again over the past 15 years, there was a fall in the share of national income going to labor.

The labor share is the fraction of total income going to wages.

Another useful accounting identity is that the **labor share** is the fraction of total income that comes as wages and salaries. In the simplest story, all income is either labor income or capital income. But we can talk about the labor share even if there are other kinds of income. It simply means the fraction of total income that is received by labor. We can write:

Labor share. The fraction of output going to workers, calculated as total wages and salaries plus benefits divided by total income. Also called the wage share.

$$\textit{labor share} = \frac{\textit{total wages}}{\textit{output}} \tag{5}$$

Here wages refers to all income that people receive as compensation for work – including both wages and salaries and noncash benefits like employers' health insurance contributions. As with productivity, we can apply this identity to the economy as a whole or to a particular sector or industry. Note that in the national accounts, the labor share in government and nonprofits is 100% by definition. So if we look just at the business sector, the labor share will always be somewhat lower than for the economy as a whole. Wages here include fringe benefits, like health insurance or pension contributions. In this equations, wages and output are measured in **nominal** terms. Output is normally measured by GDP, although we could use an alternative measure like GNP.

Nominal. Measured in units of money, not adjusted for inflation. Prices and many other numbers in economics are normally measured in money. If we try to adjust a number for changes in the value of money, that gives us a "real" figure. If we don't make any such adjustment but simply use the money value as is, that is a nominal figure.

As long as we use the same **price index**, we could just as well divide real output by real wages. As a reminder, real wages are simply equal to nominal wages divided by a price index. The change in real wages is approximately equal to the change in nominal wages minus inflation.

$\%\Delta \textit{real wage} \approx \%\Delta \textit{nominal wage} - \textit{inflation}$

Price index. A measure of the average price of goods and services at a given time and place. If the price index is 1 percent higher in one year than another, that means the price of the "typical" good is 1 percent higher in that year. Since prices don't all change together, a given price index is defined only for a particular basket of goods.

The division of output into the different kinds of income generated in production is called the division into **factor shares**, or the **functional distribution** of income. In modern economies, labor and capital are the only two important factors of production, so the functional distribution of income just means the division of total income between wages and profits. An increase in the labor share implies a decrease in the capital share, and vice versa.

Factors. Labor, capital and others who must be paid for their contributions to production.

Functional distribution. The division of total income into payments to labor (wags and benefit) and payments to capital-owners (profits). In principle the functional distribution includes payments to other factors of production, but in modern economies labor and capital are by far the most important. The other way of thinking about income distribution is the personal distribution

The change in the average nominal wage is equal to the change in average productivity, plus the inflation rate, plus the change in the wage share.

Equation 5 is an accounting identity. We also know that total wages equals the average wage times total employment, the real wage equals the nominal wage divided by the price level, and the change in the price level is inflation.

Substituting these identities into Equation 5 and applying the linear approximation gives us[2]

[2] Exactly how this is derived is shown in the box nearby.

$$\%\Delta \textit{nominal wage} \approx \%\Delta \textit{productivity} + \%\Delta \textit{prices} + \%\Delta \textit{wage share}$$

In other words, the percentage increase in the average nominal wage must be equal to the sum of the percentage increases of labor productivity, the price level, and the wage share. And since inflation is just the percentage change in the price level, we can rewrite this as:

$$\%\Delta \textit{nominal wage} \approx \%\Delta \textit{productivity} + \textit{inflation} + \%\Delta \textit{wage share} \qquad (6)$$

To think about what this means, imagine a business that for whatever reason decides to increase wages. What can happen as a result? It might be that profits will fall – that is a rise in the labor share. It might be that it will increase its prices – over the economy as a whole, that is the same as inflation. Or it might be that the higher wages will cause the business to become more productive, perhaps because workers will be more concerned about losing such a good job or because they will feel a greater sense of loyalty. Any of these outcomes are possible. But what we know for sure is that a one percent increase in wages *must* result in some combination of a higher wage share, higher prices, and/or higher productivity, that add up to one percent. This is true both at the level of an individual business and for the economy as a whole.

Equation 6 lets us think systematically about many things that happen in the economy. For instance, what happens if labor productivity grows more rapidly, while nominal wage growth is unchanged? The equation says that in this case either inflation or the wage share must fall. The fact that more rapid productivity growth is *deflationary* – tends to lead to lower prices – is not obvious, but the equation makes it clear.

Macroeconomic theory often assumes that the wage share is fixed. This implies that the increase in real wages (the increase in nominal wages less inflation) must be just equal to the growth of productivity. But in the real world, this is often not the case. Over the past 15 years, nominal wages have increased by an average of 2.9 percent a year, inflation has averaged 2.4 percent a year, and productivity has increased by 1.1 percent per year. This means that the labor share has fallen by a bit over half a percent per year. ($2.9 - 2.4 - 1.1 = -0.6$.) While this might not seem like much, over 15 years that adds up to a 9 percent total decrease – a substantial fall in the share of income going to workers.[3]

[3] This is a *9 percent decrease*, not a *9 percentage point* decrease.

We can turn Equation 6 around and ask what happens when productivity increases.

$$\%\Delta productivity \approx \%\Delta nominal\ wage - \%\Delta prices - \%\Delta wage\ share$$

Again, think of an individual business: Rising labor productivity means they are now able to produce the same quantity of goods with fewer workers. Let's say the number of workers required per unit of output has fallen by 10 percent. What happens? Either the remaining, more productive workers can each be paid 10 percent more; or the company can cut its prices by 10 percent; or the gains from increased productivity can go to higher profits (i.e. a lower labor share). Again, the accounting identity doesn't tell us which of these outcomes will happen. But it does tell us that one of them – or some combination – *must* happen whenever productivity rises.

Note: Deriving Equation 6

We know that the wage share is equal to total nominal wages divided by nominal output. This implies that

$$total\ nominal\ wages = nominal\ output * wage\ share$$

We know that nominal output is equal to real output times the price level. So substitute that in:

$$total\ nominal\ wages = real\ output * price\ level * wage\ share$$

We know that real output is equal to employment times productivity, so substitute that in:

$$total\ nominal\ wages = productivity * employment * price\ level * wage\ share$$

Finally, we know that the average wage is equal to total wages divided by total employment, so we divide both sides by employment to get:

$$nominal\ wage = productivity * price\ level * wage\ share$$

And now we apply the linear approximation of Equation 1, and that gets us Equation 6.

In an equation describing the economy, we think of some variables as fixed, some as changing on their own, and some as changing in response to the other variables.

Economists often describe relationships they think exist in the real world in terms of equations. A set of one or more equations that is intended to give a self-contained description of some outcome or phenomenon is called a **model**. Equations in a model contain variables and **parameters**. The variables have values that correspond to some real-world phenomenon – in macroeconomics normally some **aggregate** like the unemployment rate, the interest rate or the the trade balance. (In microeconomics the variables are most often the price or quantity of some particular good traded in a market.) The parameters have values that don't represent anything in themselves, but describe the relationships between the variables. Parameter values are often estimated through statistical analysis.

Model. A set of equations used by economists and others to describe or predict the behavior of some system in the real world.

Parameters. Numbers in an equation that describe the relationships between the variables.

Aggregate. A variable measured at the level of the economy as a whole. Common aggregates include GDP, the consumer price index (CPI), and the unemployment rate.

One thing that makes an equation an economic model is that the variables are classified as **exogenous** or **endogenous**. Exogenous variables are considered to be determined outside the model. They may be fixed or they may change, but in either case we have to already know their value, or to assume some value for them. The model doesn't tell us what their value should be. The endogenous variables are the ones whose values are determined within the model. If we know the values of the parameters and the exogenous variables, the model will predict values for the endogenous variables.

Exogenous. Determined outside the model. Variables that a model does not try to explain, but simply takes as given.

Endogenous. A variable that is determined by other variables, as opposed to an exogenous variable that is fixed by policy or by nature.

If a model has exactly as many endogenous variables as it has equations, we will be able to calculate values for all the endogenous variables. In this case, we say the model is **closed**. If the model has more endogenous variables than equations, the model is *underdetermined* – there will be an infinite set of values for the endogenous variable(s) that satisfy the equations, so the model gives us no prediction for which of them to accept. Underdetermined models can still be useful, since they tell us that only certain combinations of the endogenous variables are possible. If the model has more equations than endogenous variables, it is *overdetermined* – there will be no values of the endogenous variable(s) that satisfy the equations. If a model is overdetermined, that always means that an error was made in setting it up, which needs to be fixed before it can be used.

Closed model. A model is said to be closed when it has exactly as many equations as variables. In that case, it can be used to calculate values for all the endogenous variables. A model with more equations than variables is overdetermined, while a model with more variables than equations is underdetermined.

The most common way to use an economic model is to imagine an exogenous variable changing, and then ask what happens to the endogenous variable(s) as a result. In effect, the exogenous variable is the dog wagging, and the endogenous variable(s) the tail getting wagged. Different ideas about which variable should be thought of as the dog and which should be thought of as the tail is the basis of many disagreements in economic theory. Note, though, that exogenous and endogenous are descriptions of how a variable is being used in a model, not a statement about the real world. In reality, everything is linked to everything else and we can't separate exogenous causes from endogenous effects. We just choose to treat certain things as exogenous when we think that will be helpful in answering certain questions - when we want to treat one feature of reality as fixed so we can focus on changes in another feature.

In some models, the division of the variables into exogenous and endogenous is a fixed feature of the model. For example, we always think of **Okun's law** as describing how unemployment changes in response to output growth. So in the simple model represented by Okun's law, output growth is always exogenous and the change in unemployment is always endogenous. Other models describe a relationship between some variables but leave it up to the person using the model to decide which variables to treat as exogenous and which ones to treat as endogenous. The choice of which variables to make exogenous and which endogenous is sometimes described as the choice of **closure**. Another way of thinking of this is, what adjusts? What features of the world represented in the model do we think have to change to accommodate changes in the others?

Okun's law. An empirical law in economics that says the change in unemployment ΔU is connected to the real growth of output g by a relationship of the form $\Delta U = -a(g - b)$. For the US, a is around 0.6 and b is around 2.

Model closure. The choice of which variables in a model to treat as exogenous (determined from outside) and which to treat as endogenous (determined within the model).

Whenever you use an economic model, you have to decide which variables are exogenous and which are endogenous. In whatever story you are telling with the model, the cause has to be represented by an exogenous variable, and the effect by one or more endogenous variables.

The effect of productivity growth on employment and output depends on whether the economy is demand-constrained or supply-constrained.

We know that employment, output and productivity are linked by the identity:

$$\%\Delta productivity \approx \%\Delta output - \%\Delta employment$$

The percentage change in productivity is always approximately equal to the percentage change in output minus the percentage change in employment. For example, in the US in recent years, real output has been increasing by about 2% a year, with employment and productivity each increasing by about 1% a year. What if labor productivity in the US increases by 3% next year (a high rate by historical standards)? That would be consistent with the same 1% increase in employment and a 4% increase in output. It would also be consistent with the same 1% increase in output change and a 2% decline in employment. Which of these possibilities is more likely?

The answer depends on whether we think the economy is **demand-constrained** or **supply-constrained**. If the economy is demand-constrained, that means that output is limited by how much people (including businesses and governments as well as households) want to buy. Demand-constrained businesses could make more stuff, but there is no reason to since they couldn't sell it. In a demand-constrained economy, faster productivity growth is likely to reduce employment rather than raise output. Since faster productivity growth, by itself, doesn't increase how much anyone wants to spend, there is no reason for businesses to produce more. So they are more likely top respond to productivity improvements by producing the same output with less labor. In a supply-constrained economy, output is limited by the real resources available, primarily technology and labor. In a supply-constrained economy, productivity increases will allow businesses to produce more output. Suppose the economy is already at full employment - everyone who wants to work, has a job. (Or more realistically, unemployment is as low as policymakers believe it can safely get.) In this case, businesses in the aggregate can't raise output by hiring more labor. (Of course individual businesses can hire workers away from each other.) So faster productivity growth will allow greater output with the same amount of labor.

Demand constrained economy. An economy where output is limited by how much people want to buy, rather than by the productive capacity of business. Most economists believe that modern economies are demand-constrained in the short run by supply-constrained when we are considering periods of mroe than a couple years.

Supply-constrained economy. An economy where output is limited by real resources – labor, technology, etc. – rather than by how much people want to buy,. Most economists believe that modern economies are demand-constrained in the short run by supply-constrained when we are considering periods of more than a couple years.

In the short run, we believe the economy is almost always demand-constrained. But we also don't think that there are large *exogenous* changes in productivity growth in the short run. (In other words, if labor productivity accelerates or slows down a lot between one year and the next, that is more likely to be in response to changes in output and/or employment, rather than the cause of them.) Exogenous changes in productivity growth, due to technology or other factors, are likely to come over longer periods of several years or even decades - it takes time for new products and techniques to be developed, and even longer for them to be widely adopted. Over these longer periods most economists believe that it makes more sense to think of economies as supply-constrained. So faster productivity growth will translate into higher output, rather than lower employment.

Faster productivity growth can lead to faster wage growth, lower inflation, or a lower labor share.

We know that productivity growth, nominal wage growth, inflation and the labor share are linked by the identity

$$\%\Delta productivity \approx \%\Delta nominal\ wage - \%\Delta prices - \%\Delta wage\ share$$

The percentage change in productivity is always approximately equal to the percentage change in nominal wages - the percentage change in prices (or the inflation rate) minus the percentage change in the wage share. Since the change in nominal wages minus inflation is the same as the change in the real wage, we can also say that productivity growth is always approximately equal to real wage growth minus the change in the labor share. So an exogenous **acceleration** in productivity growth could mean higher wages; or it could mean lower prices; or it could mean a lower wage share.

Acceleration. An increase in the growth rate or rate of change of a variable. With many macroeconomic aggregates, such as output or the price level, we are more interested in the rate of change than the level. So we often want to talk about changes in the rate of change. When the growth rate or rate of change of a variable increases (or the decline in a variable slows down) we call that an acceleration. When the growth rate slows (or the decline speeds up) we call that a deceleration.

Think about an individual business, let's say a road contractor. Suppose a new machine comes on the market that allows the same number of workers to put down a longer road in the same time (or equivalently, the same road with fewer workers). The contractor could increase wages by the same proportion as output per worker has increased - for example, if each worker can now produce 50% more in the same period, the contractor could increase their hourly wage by 50%. This would leave the contractors's costs per mile unchanged. Or, they could cut their prices – if it takes only two-thirds as much labor to produce each mile of road (and there are no other costs), it could cut the price it charges per mile to two-thirds of the old price. If it doesn't do either of these things – if it leaves wages and prices the same, even though it now takes fewer workers to produce the same length of road – then by definition the share of sales going to profits must increase. If costs fall but the business charges just as much, then profits – which are just what is left over after costs – must increase.

We might expect that business will cut prices if it is under pressure from competitors. It will raise wages if it has to compete to hire workers, or if workers are organized in a union or otherwise in a position to demand higher wages. If neither of these applies, then most businesses will probably prefer to take at least a large part of productivity gains in the form of a higher profit share (or equivalently, a lower labor share.)

At the level of the economy as a whole, we often think that an increase in productivity will lead to higher wages. It is certainly true that higher productivity makes it possible for real wages to rise, without cutting into profits. But this won't necessarily happen on its own - workers have to be in a position to demand higher wages. Between 1945 and 1980 or so, real wages and productivity did mostly move together. But in the past three decades, the relationship has been less consistent. Except during the late 1990s, faster productivity growth has not generally been associated with faster wage growth.

If competition between businesses is intense, then if productivity gains are not passed on to wages, they are most likely to show up as lower prices. Under these conditions, faster productivity growth is **deflationary** – it tends to cause prices to fall. For a country with a great deal of foreign trade, this may be considered desirable, since it makes its products more **competitive** in foreign markets. But in other cases, this can be a problem, since falling prices can be economically destructive, in part because they make it more difficult for debtors to service their debts.

Deflation. Negative inflation, or a decline in the price level.

Competitiveness. The cost of producing a good in one country compared with the cost of producing similar goods elsewhere. A country will be more competitive if its costs – especially wages – are lower than elsewhere, or if its industries are more productive.

If workers are not able to demand higher wages and businesses are not forced by competition to pass productivity gains on. as lower prices, then faster productivity growth will imply higher profit margins and a lower labor share.

When nominal wages rise, there must be an equal rise in prices, productivity or the wage share, or some combination of the three.

Another way of writing the same identity is:

$$\%\Delta \textit{nominal wage} \approx \%\Delta \textit{productivity} + \%\Delta \textit{prices} + \%\Delta \textit{wage share}$$

Faster growth in nominal wages implies some mix of faster productivity growth, higher inflation, or an increase in the share of output going to labor. Suppose that business finds itself forced to pay higher wages - perhaps because of a strong labor market that forces businesses to compete for scarce workers, or perhaps because of a higher minimum wage or some similar legal change. The higher wages may force businesses to find ways to raise productivity, perhaps by substituting machines for workers, or simply by organizing work more efficiently. Higher wages may also encourage greater effort from workers, and reduce costly turnover – this is called an **efficiency wage** effect. Finally, higher wages may simply force less productive firms out of business, since they can't afford to pay the new higher wage. If the least productive firms exit, leaving only the more productive ones, this will raise average labor productivity. Some combination of these three effects may result in some or all of an exogenous wage increase translating into higher productivity, leaving businesses' costs unchanged.

Efficiency wage. Efficiency wage refers to the idea that paying workers more than the going wage may increase effort and reduce turnover, raising productivity.

If wages increase and productivity does not, or if productivity increases but by less than the wage increase, then costs per unit of output will rise. Wages (including benefits) per unit of output are called **unit labor costs**. If businesses are able to pass higher costs on in the form of higher prices, then an exogenous increase in unit labor

Unit labor costs. Wages and salaries per unit of output. Unit labor costs are normally measured in nominal terms; real unit labor costs are the same as the labor share of output.

costs will produce an equal rise in the price level. Recall that the change in real wages is equal to the change in nominal wages minus inflation. So if rising unit labor costs are passed on to prices, then an increase in nominal wages will increase real wages only insofar as it leads to higher productivity.

Finally, if an exogenous increase nominal wages leads to neither higher productivity nor higher prices, then it has increased businesses' costs without increasing their sales. This means that a greater share of output is going to labor, and there is less of a surplus left for the owners.

The traditional view held by many policymakers is that exogenous changes in wages are mostly or entirely passed through to inflation. But there is not as much agreement about this as there used to be.

In most economics textbooks and many policy discussions, it is assumed that productivity increases are exogenous – they are the result of technological or other "real" developments that have nothing to do with wages, the labor market, or the state of aggregate demand. It is also often assumed that the labor share changes very little over time. If both these assumptions are true, then an exogenous change in wage growth has to lead to an equal change in inflation. Anything that causes wage growth to accelerate by one point, for example, will also cause inflation to increase by one point. This means that raising nominal wages won't do anything to raise real wages; to raise real wages, some way has to be found to speed up productivity growth.

The view that productivity growth is exogenous and *factor shares* are fixed is not as widely held as it used to be. In the decades after World War II, the division of output between labor and capital was quite stable, but in more recent decades it is has become more variable, with the labor share generally falling but sometimes rising. The fact that the periods of rising labor share have been periods of low unemployment and strong labor markets suggests that macroeconomic conditions may be important for changes in factor shares. The work of Thomas Piketty and other economists has also been important in focusing attention on changes in the labor share.[4]

[4] For an introduction to Piketty's work, see Paul Krugman's review of his book *Capital in the 21st Century*, "Why We're in a New Gilded Age", in the *New York Review of Books*.

At the same time, there is increasing evidence that weak labor markets can contribute to slower growth in productivity, while strong labor markets can lead to faster growth. Both for individual businesses and at the level of the economy as a whole, rising wages seem to be an important spur to productivity improvements. The rapid productivity growth in the late 1990s, in this view, benefited from the low unemployment of that period, while the weak labor markets of the past decade may be one reason that productivity growth has been slow compared with the past. If the currently low unemployment rate (3.7 percent as of September 2018) persists for several years, we will get an important test of this theory.[5]

One reason that higher wages might lead to higher prices is if businesses engage in markup pricing.

Why might an exogenous increase in wages lead to an equal increase in prices? One answer is that many businesses set prices as a **markup** over the marginal cost - the cost of producing one additional unit. For example, a business might aim for a markup of 20 percent; in that case, if a product that cost $100 to produce they they would charge $120 for it. Markup pricing applies to the costs of production, not to wages specifically; but for most businesses wages are the single largest component of costs, and for the economy as a whole, the costs of many non-labor inputs resolve into wages. (For example, when a business pays for electricity, part of the cost reflects wages paid by the utility that produces it.) So if markup pricing is widespread, it makes sense to expect that an exogenous increase in wages would translate into higher prices. Of course not all costs are wages, even indirectly - for example, some raw materials are imported. So even if every business used strict markup pricing, we would not expect 100 percent of changes in wages to be passed through to prices. But as a rough first approximation it might not be far off.

Of course there are reasons why businesses might not practice strict markup pricing. If their competitors don't face the same cost increase, they might worry about losing market share. Changing prices frequently may be costly or inconvenient. For some products – health care services, say – it may be be difficult or impossible to determine what marginal cost is. But for many businesses, the idea that prices are set as a markup over costs is probably reasonable.

Mathematically, it is possible to show that the profit-maximizing markup is negatively related to the **price elasticity** of demand - what fraction of sales a business will lose when it raises its prices. Specifically, if P is the price, C is the marginal cost, and e is the price-elasticity of demand, then the the profit-maximizing price is given by:

$$P = C \cdot 1/(1-1/e) \tag{7}$$

For example, suppose Apple believes that if it raises the price of an iPhone by 10%, it will lose 20% of its sales. That implies a price elasticity of 2. Equation 7 then suggests that they should charge a price of $1/(1-1/2) = 1/(1/2) = 2$ times cost, or another words a markup of 100

[5] Some reasons for thinking that weak demand and high unemployment could lead to slower productivity growth are discussed in Section 3 of J. W. Mason, "What Recovery? The Case for Continued Expansionary Policy at the Fed".

Markup pricing. The idea that businesses set prices at a fixed percent above the cost of production.

Price elasticity. How much sales of something change in response to a given change in price. For example, if a one percent increase in price leads to a one percent fall in sales, then price elasticity is one. If a one percent increase in price leads to a 5 percent fall in sales, the price elasticity is 5. If a two percent increase in price leads to a one percent fall in sales, price elasticity is 0.5. Generally, the easier it is to find a substitute for a good – either the same good from a different supplier, or something else that meets the same need – the more price elastic demand for it will be.

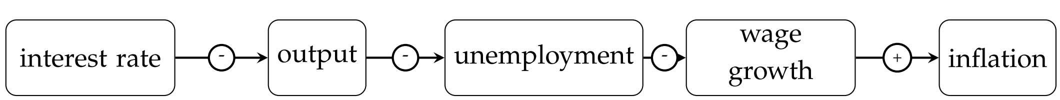
percent. (This would be a large markup, but a company like Apple has a lot of market power.) A company that thought it would lose 90% of its sales if it raised prices by 10% would markup its prices over costs by a bit over 12 percent. ($1/(1-1/9) = 1/(8/9) = 1.125$.)

In real life, of course, things are not so simple. A business may have other goals besides maximizing short-run profits - for instance, building up market share. And the price elasticity of demand may not be clear, especially in an **oligopolistic** market where the effect of a price change will depend on how competitors respond. Still, it is reasonable to suppose that markups will be low for products with price-elastic demand (where there are close substitutes available) since if a business tried to increase their prices much above costs they would quickly lose most of their sales. More broadly, stable markups are the most straightforward way to explain why increases in wages would normally lead to higher prices.

Oligopoly. A market where there are a small number producers selling equivalent products.

The relationship between wage growth, productivity, inflation and the wage share is important for macroeconomic policy.

Conventional macroeconomic policymaking relies on the idea that changes in wages and changes in prices are closely linked. Traditionally, macroeconomic policymakers have assumed that productivity growth varies for reasons outside their control, due to factors technological change or the discovery or exhaustion of natural resources. And they have assumed that factor shares do not change very much. So in order to achieve a 2 percent target for inflation, it is necessary for wages to grow at the exogenously determined rate of productivity change, plus 2 percent. Wage growth in turn is assumed to depend on the unemployment rate, with wages growing more slowly when unemployment is high and more rapidly when unemployment is low. The unemployment rate is assumed to depend on the growth of output. And output growth is influenced by the tools of policy - meaning the interest rate for the central bank. The logic is shown below:



In this diagram, higher interest rates lead to lower output. Lower output leads to a higher unemployment rate. Higher unemployment leads to slower wage growth. (These first three relationships are negative - an increase in the first variable leads to a decrease in the second – so they are shown with a small minus sign.) And slower wage growth leads to lower inflation.

In this framework, the central task of policymakers is to guess what medium-term productivity growth rate is likely to be, and then try to raise the unemployment rate to get wage growth to a pace two points above that. If productivity growth is high, thanks perhaps to a major new innovation, then the economy can sustain faster growth and lower unemployment; if productivity growth is expected to be slower, then growth must be held down and unemployment kept up. In any case, the goal is to keep inflation around 2 percent.

For example, around 1995 unemployment in the US had fallen to historically low levels, and many people at the Federal Reserve were worried that wage growth would soon accelerate, leading to an undesirable rise in inflation. But the head of the Fed at that time, Alan Greenspan, believed that thanks to the new technologies associated with the tech revolution – faster and cheaper computers, the internet and so on – there was a good chance that productivity growth would accelerate as well. Under those conditions, low unemployment and faster wage growth would be consistent with stable inflation. So he resisted calls for the Fed to shift toward more **contractionary** policy, and allowed unemployment to stay low. As it turned out, in the late 1990s the US did experience a period of rapid productivity growth, and rising wages without higher inflation. This is widely considered a success story for macroeconomic policy.[6]

Contractionary. Has as its intended or primary effect a reduction in output.

[6] Imagine that productivity growth was 2%, and nominal wages were rising by 4%. Then if factor shares were fixed, inflation would be 2%. Now suppose you expected productivity growth to rise to 3% per year. If wage growth stayed at 4% and factor shares were still fixed, that would imply a fall in inflation to 1% ($3\% = 4\% - 1\% - 0\%$), below the Fed's target. To keep inflation at 2%, wage growth must rise from 4% to $3\% + 2\% + 0\% = 5\%$. For wage growth to accelerate, unemployment must go lower.

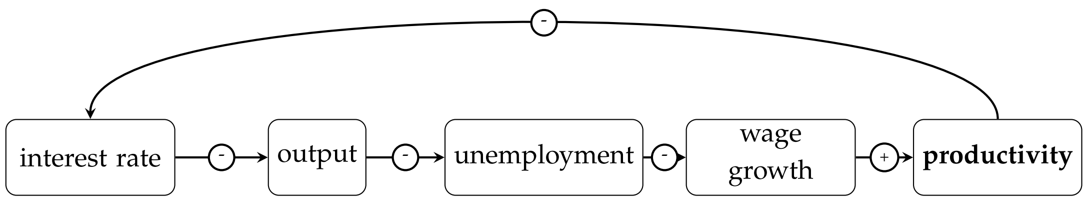
This approach to policy will produce unintended effects if wage changes are linked to productivity growth and/or factor shares.If changes in wage growth mainly affect the labor share, then *contractionary* policy that is intended to hold down inflation, will instead redistribute income from labor to capital. **Expansionary** policy will similarly redistribute income from capital to labor. In this view, the fall in the wage share of US GDP over the past generation might reflect excessively contractionary monetary policy. If changes in wage growth lead to changes in productivity, on the other hand, macroeconomic policy will be, without realizing it, reacting to its own effects. If expansionary policy allows unemployment to fall and wage growth to accelerate, then productivity growth will also accelerate, justifying the lower wages. On the other hand, if slower wage growth leads to slower productivity growth, then the central bank will believe that sustainable wage growth is lower, and seek to lock the new lower wage growth in place. In this case, rather than bring wage growth into line to exogenous changes in productivity growth, policymakers will be creating whatever productivity growth corresponds to the wage growth they are choosing. This implies a **positive feedback** between policy and productivity - strong productivity growth en-

Expansionary. Has as its intended or primary effect an increase in output.

Feedback. When a change in one variable leads to further changes that eventually produce a further change in the original variable, that is a feedback loop. In a positive feedback, the effect of the further changes is to push the original variable further in the direction it moved initially. In a negative feedback, the effect of the further changes is to weaken or cancel out the original change.

courage more expansionary policy, leading to low unemployment, faster wage growth and further productivity growth. Weak productivity growth will have the opposite effects. Some economists argue that this is at least part of what happened in the 1990s – Greenspan's decision to keep rates low wasn't just based on an accurate forecast of accelerating productivity, but actually helped produce that acceleration.

This positive feedback loop is shown below:



The logic is the same as in the previous diagram, except that now faster wage growth leads to faster productivity growth, rather than higher inflation. This higher productivity growth is then taken as a signal by the central bank that they can safely lower interest rates. On the other hand, if wage growth slows for any reason, productivity growth will slow, and the central bank will take this as a signal that they need to keep wage growth slower in the future. If you walk through the diagram, you can see that a rise in wages will lead to a further rise in wages, and a fall in wages will lead to a further fall.

In recent years, some economists have been concerned that the weak productivity growth of the past decade may reflect weak demand, rather than a lack of new technologies. If that is the case, then if the Fed sees slow productivity growth as a reason to keep wage growth down, they will actually be creating the problem they think they are responding to. Several economists at the New York Federal Reserve Bank wrote recently: "One interpretation of the listless recovery is that recessions inflict damage on an economy's productive capacity. For example, extended periods of high unemployment can lead to skill losses among workers, reducing human capital and lowering future output. ... Monetary policy rules that fare well in normal times can lead to pathological outcomes when faced with large shocks." If the authorities treat the damage done by the recession as a sign that the economy's capacity is simply lower than believed, they will refuse to allow the strong expansion that could reverse the damage. In effect, a belief that hysteresis just reflects the "new normal" can be self-confirming. Instead, faced with hysteresis, the central bank should aim for a period of exceptionally strong demand, with levels of unemployment and wage growth that would normally be considered inflationary: "Just as recessions damage potential output, booms can repair it."[7]

[7] Acharya, Bengui, Dogra and Wee, 2016. "Escaping Unemployment Traps".

Other prominent economists have been concerned about the possible link from wage growth to the wage share. Jon Faust, a senior economist at the Federal Reserve, suggests that following years of a declining labor share, there may be good reasons to expect "a return of labor's share to something closer to prior levels. In accounting terms, the most obvious way for this to happen is for nominal wages to grow faster than the sum of inflation and productivity growth. ... In the real world that we might soon face, policymakers may have to take a position on whether the rising wages are a natural part of a secular – and to many, a desirable – re-balancing in labor's share or instead are a sign of cyclical overheating. A sharp monetary policy response to what happens to be beneficial secular dynamics could have the undesired consequence of slowing the economy, raising unemployment, and delaying the secular rebound in labor's share."[8] In other words, if policymakers see any acceleration of wage growth as sure to lead to higher inflation, they may end up preventing any increase in the share of income going to labor rather than capital.

[8] Faust and Leeper 2015, "The Myth of Normal: The Bumpy Story of Inflation and Monetary Policy".

Most orthodox economists probably still believe that in normal times, the main link is between wage growth and inflation. But there is more interest than there used to be in the other possible links. The most important thing to know is that real-world discussions of wage growth and inflation take place in terms of the accounting identity described in this section.