

cant increase in price, by a factor of about 6, and this combined with the increase in real value-added to generate the gain in value-added *share* for that industry.

Figure 7.3 also shows that the price level of manufacturing rose, but not by nearly as much as that in the other industries. The combination of a mild increase in real value-added output and a mild increase in price combined to make the share of manufacturing in GDP fall over time. It was not that we produced fewer manufactured goods but that relative to other industries, production did not go up by as much as it did in those industries, and just as important, manufactured goods got cheaper over time relative to services.

As we saw already, these shifts had a significant effect on aggregate productivity growth. What I'm going to show next is that the shifts we just saw in figures 7.2 and 7.3 are what we'd expect to see as an economy becomes rich, and they are the consequences of success, not failure.

BAUMOL'S COST DISEASE



To explain why the shift into services occurred, why that shift was associated with lower productivity growth, and why it represents a success, I'm going to talk about the work of an economist named William Baumol.

Over the course of nearly seven decades of scholarship, Baumol contributed to many areas of economics, but perhaps the ideas he is most known for relate to the fundamental differences between service production and goods production. The origin of these ideas can be traced to two articles he published back in the mid-1960s. Both articles are what you might call "old-style" economics, explaining concepts with words rather than a series of equations (OK, there were some equations but not many). It makes the articles intelligible to almost anyone with some rudimentary economics training, although at times the lingo can get a little thick as well. Baumol stayed active until very recently (he passed away in 2017), publishing several more-accessible versions of his ideas for the public.

THE IMPORTANCE OF TIME

In his two original articles Baumol laid out his ideas about what makes services so different from goods. The quotes that follow are all from his 1967 article, which I find most relevant for our purposes here. To start, Baumol divided economic activity up in the following way:

The basic source of differentiation resides in the role played by labor in the activity. In some cases labor is primarily an instrument—an incidental requisite for the attainment of the final product, while in other fields of endeavor, for all practical purposes the labor is itself the end product.

For the first case, labor as an instrument, you can read “goods.” Baumol uses the example of an air conditioner. Just by looking at an AC unit, or using it, there is no way to assess how much labor went into producing it—nor do we really care. The labor that went into producing it is incidental from the perspective of a consumer of the product. The same goes for cars, houses, refrigerators, laptops, and smartphones.

Contrast that to the following:

On the other hand there are a number of services in which the labor is an end in itself, in which quality is judged directly in terms of amount of labor. Teaching is a clear-cut example. . . . An even more extreme example is one I have offered in another context: live performance. A half hour horn quintet calls for the expenditure of 2½ man hours in its performance, and any attempt to increase productivity here is likely to be viewed with concern by critics and audience alike.

For most services, labor is the very essence of the product. If you go hear the horn quintet, as Baumol suggests, then you are buying the half hour of each player’s time. At many restaurants, you are purchasing not just the food, but the time and attention of a waiter. For the service industries that are growing over time, like education, health care, and professional services, you are almost always purchasing people’s time or attention as opposed to any tangible good.

If you go to the doctor, you want the doctor to spend some time with you, explaining why your back hurts and what your treatment options are. You want and need the doctor’s attention to your condition, and there is no substitute for that expert time and attention. If you go to an hour-long yoga class because your doctor told you it would help your back, then you want sixty minutes of yoga instruc-

tion, and the instructor’s attention on your form so that you don’t hurt yourself again and can get the most relief from your back pain. If yoga doesn’t work and the doctor screws up surgery on your back, then you’ll call a lawyer to sue, and what you are buying is the lawyer’s time and attention, which will be apparent because the lawyer will bill you in great detail for time, not for a specific product. Compare that to a good like an AC unit. The people working at the AC factory are probably lovely, but I don’t want, or need, to spend any time with them every time I try to cool down my house.

For Baumol, the instrumental role of labor explains why productivity is high and, more important, can grow very quickly for the first kind of production (goods) but is low and grows slowly in the second kind (services). In producing goods, you can “do more with less,” meaning that a company could invent ways to get more AC units or cars or laptops out of its existing labor force, or it could invent ways to get the same number of units by using a smaller labor force. The time spent by the workers isn’t what matters to the consumer, so firms can shed workers or raise output per worker as much as possible without damaging the demand for their goods. They have every incentive to decrease the labor used, as doing so lowers their own costs.

But for services, you cannot “do more with less.” No one wants to see the half-hour horn quintet played in only twelve minutes, or hear the quintet play two separate pieces at the same time for the whole half hour. You can’t give a one-hour yoga class in less than sixty minutes. You can’t achieve the same level of attention and service at a nice restaurant by halving the number of waiters. It takes a half hour to get my eyes checked by my optometrist; getting out of there in fifteen minutes would mean she checked only one eye. The time and attention of the workers in these industries are what we as consumers value, so if firms cut back on that time or attention, then demand for their service will suffer.

The difference in the role of labor in goods and services, and the implication of this for productivity growth, is not some absolute law of the universe. Yes, there are ways to increase productivity in some services by cutting back the labor used without changing demand much. A yoga instructor could add more students to the class. Online

courses allow university professors to teach more people in the same amount of time. Remote medicine can free up a doctor's time to meet with more patients in a given day. But note that even these examples will run into issues at some point. Too many people in a yoga class and the instructor cannot observe and correct everyone's form. Online courses have a similar problem, in that the professor cannot offer individual feedback to every single student. Remote medicine works to the extent that a doctor does not need to make an actual physical examination. The scope of productivity improvement in many services is constrained by limits on the actual time and attention span of the providers.

This was Baumol's first key insight. Compared to goods production, the productivity growth of services is going to be relatively low. And we saw exactly that in the data in the previous chapter on the productivity growth rate across different industries in the past fifteen years.

THE COST DISEASE

The difference in productivity growth between goods and services led to Baumol's second insight, which he named the *cost disease* of services. Every time a firm gets more productive—meaning that it can produce more output with fewer inputs (e.g., the time of its workers)—its costs go down. Because goods-producing firms can achieve higher productivity growth than service-producing firms, the cost of producing goods falls faster than the cost of producing services. This means that the relative cost of goods keeps getting smaller and smaller as compared to services. That means that the relative cost of services must be getting higher over time as compared to goods.

In a market economy, it should be the case that prices and costs move together. So Baumol's cost disease of services should show up as the relative price of goods getting lower and lower when compared to that of services. And this is in fact what we see in the data.

Figure 8.1 plots a price index for each of eight types of products. There are three products I would consider clear services: higher education, health care, and food service. There are three that I would

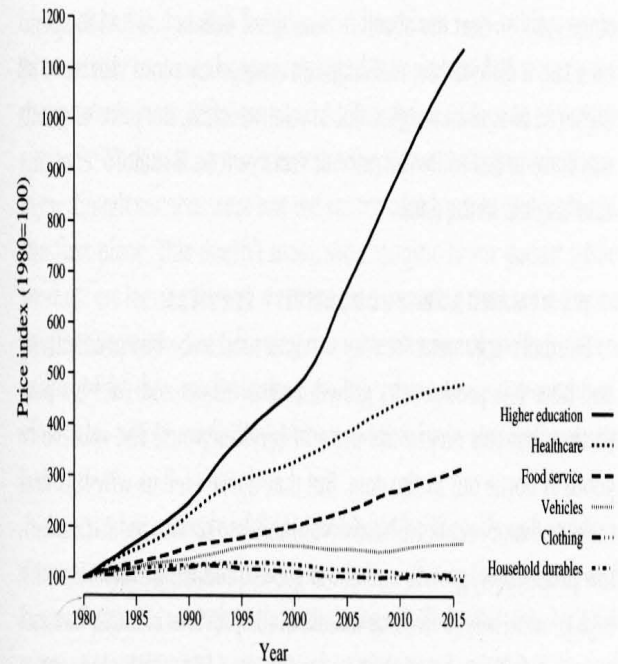


Figure 8.1. Price indexes by type of product

Note: Data is from the Bureau of Economic Analysis personal consumption expenditure table 2.4.4. The author made the calculations to index all price indexes to 100 in 1980.

consider clear goods: vehicles, clothing, and household durables (e.g., a dishwasher).

The distinct trajectories of the services and goods are quite apparent and entirely consistent with Baumol's theory. The growth of prices in higher education and health care far outstrip any other product, and even prices in food service are growing much faster than the prices of any of the goods plotted in the figure. The differences are huge, and likely something you are well aware of, in particular if you've been to the hospital or had a child go through college recently. Higher education prices went up by a factor of almost 12 in 2016 compared to 1980, while health-care costs were higher by a factor of almost 5. But even food service had prices three times higher in 2015 than in 1980. For two of the goods plotted, clothing and household durables, the data shows that their prices have remained flat over time, whereas for vehicles, there was only slight growth in their price.

Note that none of these statements is about the *absolute* cost of these goods. General inflation in prices has occurred in all products

since 1980, so that the absolute number of dollars you had to spend on a car, a dishwasher, tuition, or an emergency-room visit were all higher in 2016 than in 1980. But in relative terms, the price of goods was close to flat while the price of services rose. Baumol's "cost disease" is clear in the data.

WE DEMAND LOW-PRODUCTIVITY SERVICES

Baumol's argument for why services tend to be low productivity and have low productivity growth seems reasonable. And his prediction that this means the price of services would rise relative to goods is borne out in the data. But that doesn't tell us why it is that more and more economic activity shifted into service industries with low productivity growth and out of goods-producing industries with high productivity growth, as we saw in the previous chapter. Baumol speculated about that in the same 1967 paper. His terminology gets a little hard to understand, so I'm going to put this argument into my own words.

If the demand for services is income elastic, and the demand for goods is income inelastic, then productivity growth in any industry will translate into a greater share of spending on services. *Income elastic* means that if you currently spend half of your income on goods and half on services, then you'll spend more than half on services if I give you an extra \$100. Income inelastic is just the mirror image of this for goods. If I give you an extra \$100, then you'll spend less than half of this on goods, which has to follow if you spend more than half on new services. Over time, this means the fraction of your income that you spend on services grows as your income grows. And this is exactly what we see. Think of Bill Gates. His income is thousands of times larger than mine. Yes, he has a bigger house, more refrigerators, and nicer cars, but his spending on goods is not thousands of times more than mine. In contrast, the fraction of his income that he spends on services (e.g., travel, restaurants, legal services, business services, personal assistants) is way, way larger than mine. As people get richer, Baumol suggested, the mix of their spending changes and shifts toward services.

Because our demand for goods and services works this way, as productivity in goods keeps growing, this makes us feel richer—we can produce more with less—and we use those savings to buy more services. Workers then move into the service sector to provide them, even though services were not the source of the productivity gain in the first place. This doesn't mean we consume fewer goods; while workers are leaving those industries, the remaining workers are more productive, so we can still enjoy the same number of goods (or even more goods) than we did in the past. But we can also enjoy more services than before, even though the productivity in services may not have grown.

An extreme example of this logic would be if goods and services were perfect complements. Let's say that we all want to consume one hour of yoga for every head of lettuce we eat. To begin with, let's assume that it takes one hour of work to grow one head of lettuce, and it obviously takes one hour of work to provide one hour of yoga. If we've got twenty hours of labor to provide, we'll use it to consume ten heads of lettuce and ten hours of yoga. Now, let's say we get more productive at making lettuce (a good), so now it only takes half an hour to produce one head of lettuce. We could get twenty heads of lettuce from our ten hours of lettuce work and still consume ten hours of yoga. But we don't like the asymmetry, given our preferences. So we switch some of our hours of work over to yoga and away from lettuce production, even though lettuce production just got more productive and yoga production didn't improve at all. To get equal consumption of lettuce and yoga, we'd end up using 13.3 hours on yoga, and 6.67 hours on lettuce production (which would give us 13.3 lettuce heads). We get more of both yoga and lettuce, even though we moved labor out of the industry with productivity growth and into an industry with stagnant productivity.

This logic is why you have to be very careful about making any kind of value judgments about the continued shift into services, and so about the productivity and growth slowdowns that are a result of that. Slow productivity growth in services is due to the time- and attention-intensive nature of services and does not necessarily represent a failure

of our technological know-how or aptitude. The shift into services is a consequence of our incredible success at making goods, not a sign of some failure or problem with the economy.

Knowing what we know about Baumol's cost disease and the shift into services, we can see how plans to reinvent or restructure health care and education—two very income-elastic services—may not result in their share of economic activity declining. Let's say that we came up with a set of miracle health-care policies that not only arrested the growth in health-care costs but actually lowered health-care costs for consumers by thousands of dollars per person. To be concrete, let's say that each and every person in the country would be able to get exactly the same health care they do today, but for \$5,000 less per year, and then each person received that \$5,000 in cash.

This seems a little ridiculous, but think of the stories about \$50 emergency-room Band-Aids, or \$40 aspirins, or hundreds of thousands of dollars for a two-day stay in a hospital. It sure seems like we could, given some kind of innovation in how we deliver health care, lower costs by thousands of dollars per person. That would be a lot like a onetime boost to productivity in health care. And it seems like that would go a long way toward lowering the share of GDP that is spent on health care and might lead to workers flowing out of health care and back into higher-productivity industries.

But here is the question Baumol thought to ask: what would people spend that extra \$5,000 on? They could use it toward a new car or a major appliance, both manufactured goods. That would raise demand for those products and might pull workers into those industries, increasing the share of GDP accounted for by manufacturing. But they might spend that extra \$5,000 to take a well-deserved vacation, spending it on tourism and hospitality services. Or they might decide to spend that money sending their kids to a better (or more expensive?) school, or putting them in day care full-time rather than part-time. Or maybe they send one of their kids to community college who might not otherwise have gone. Perhaps they use the \$5,000 to send someone back to get a master's degree so he or she can get a promotion at work.

Some of the \$5,000 might even be spent on even more health services. If health care were cheaper, individuals might undertake procedures to permanently deal with chronic problems rather than only alleviating symptoms. Maybe they'd get their kids full orthodontic treatment rather than partial work that straightened only one tooth. The \$5,000 could easily be spent on visits to specialists rather than general practitioners, or to see a physical therapist, or to hire a nurse to look in on an elderly relative.

Services—and education and health services in particular—are income elastic. This means that a huge part of the money people would get back from any innovation in health care would be plowed right back into health care, education, and other services. What does that do? It increases demand for those services, which pulls more workers into those industries. The amount of spending on health care and education would still represent a significant fraction of GDP, and perhaps an even larger one. We would certainly be better off if we could raise productivity in health care, but it would likely be a onetime drop in the price of health care, and then the price would continue to grow relative to goods over time, because that's how our preferences for different types of products work.

THE LONG-RUN POWER OF PREFERENCES

In the end, that reallocation of economic activity away from goods production and into services comes down to our success. We've gotten so productive at making goods that this has freed up our money to spend on services. And because services have difficulty improving productivity as compared to goods, this means that we are merging from a fast lane into a slow lane of productivity growth. This effect has become more pronounced in the past two decades, but it is not just a recent phenomenon. Remember, Baumol wrote about these effects back in the 1960s, when goods production was still a much larger part of economic activity. Even then he could see these forces at work.

To give you some idea of how long this reallocation has been going on, we can go back in the data to just after World War II and plot the share of total personal consumption expenditures that are accounted

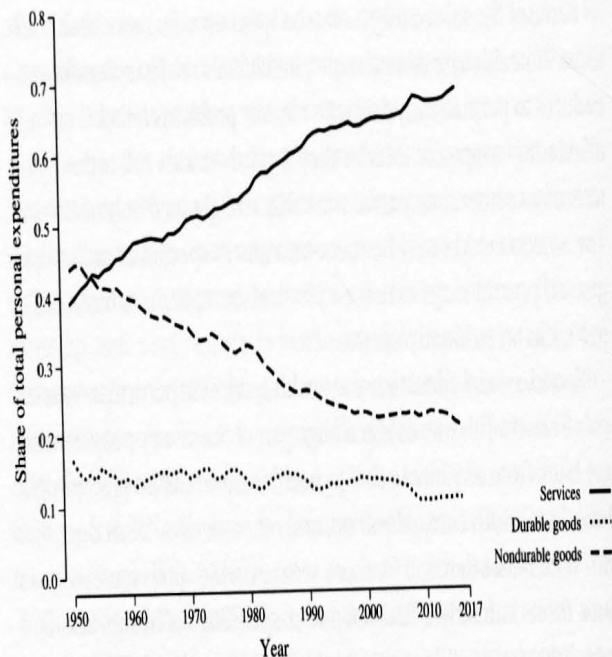


Figure 8.2. Share of goods and services in real GDP over time

Note: Data is from the Bureau of Economic Analysis personal consumption expenditure table 2.3.5. Shares of total expenditures were calculated by the author.

for by goods and services. For goods, I've divided the data up into durable (e.g., dishwashers, cars) and nondurable goods (e.g., food, clothes). Figure 8.2 plots the shares back to 1950. There are two things to note here. First, even in 1950 services accounted for a significant fraction of expenditures, about 40%. By 2017, the share of expenditures on services had climbed to around 70%.

Second, the share of spending on nondurable goods was very high in 1950, also at around 40% of total spending. That nondurable share then declined over time, halving to about 20% by 2017. In contrast, the fall in the share of spending on durable goods was not as dramatic. It started at about 16% and then declined to 11%. What drove the overall shift out of goods and toward services was a drop in our spending on day-to-day nondurable goods.

The shift into services that accounted for a portion of the growth slowdown was not a feature unique to the twenty-first century. It has been going on since 1950 (and even before) and that reinforces Baumol's point that this represents a long-standing difference in our

preferences for services (income elastic) and goods (income inelastic). What changed in 2000 was that the share of economic activity had reached such a high level that the drag on productivity growth from this shift finally became tangible.

But keep in mind that we reached this point only because of our continued success during the twentieth century in increasing productivity in goods production. The decline in the share of spending on nondurable goods represents our achievement in making those goods available at such cheap prices that we no longer have to spend a large portion of our income on them. This allowed us to consume more goods and more services, and although this in turn led to rising relative prices for services and an increase in their share of total spending, we still have increased consumption of both types of products.

ACCOUNTING FOR SUCCESS

Before moving on, it seems worth summarizing where the explanation for the growth slowdown stands at this point in the book and to remind you of where I'm headed next. The growth rate of real GDP per capita fell from 2.25% per year in the twentieth century to 1% per year in the twenty-first, a drop of 1.25 percentage points. Of that, 0.80 percentage points, and perhaps as many as 1.11 percentage points, could be chalked up to the drop in the growth rate of human capital per person. The demographic shifts behind this represented a success for two reasons: rising living standards that affected choices toward fewer kids, as well as increased opportunities and reproductive rights for women, which allowed them to have more control over their choices.

Turning to productivity growth, the previous two chapters showed that the shift away from goods production into services could account for up to 0.2 percentage points of the growth slowdown. This shift was, in turn, driven by an immense increase in material living standards. Combined, the consequences of our success can account for about at least 1 percentage point of the 1.25 percentage point drop in growth, or 80% of the entire slowdown. That leaves only a

small amount of the growth slowdown left to explain, and that small amount may be even smaller if we allow a larger role for the drop in human capital growth.

What I'm going to do for the remainder of the book is work through a number of explanations for the growth slowdown that would most likely be termed *failures* (e.g., overregulation, rising market power of firms) and show why they account for, at best, only a limited part of the slowdown. As a reminder, the argument here is not that these apparent failures are in fact successes or should be applauded or encouraged. They are failures from the perspective of people's actual well-being and economic security, but they did not play a major role in creating the growth slowdown. In the end, though, I'll come back to why that should give us some optimism about our ability to deal with these failures.

MARKET POWER AND PRODUCTIVITY

9 An economic phenomenon often put forward as an explanation for the growth slowdown is an increase in the market power of firms. This has manifested in several different trends, including a rise in the share of output that is paid out in economic profits, a matching fall in wages as a share of total output, an increased markup of prices charged by firms over their costs, and increased concentration of firms within many industries. Over the course of the next few chapters I'll provide or discuss evidence for all of these to make the case that market power did increase over the past few decades. This market power, for several reasons we'll get into, messed with the allocation of labor and capital across firms, with possible consequences for productivity growth. But it turns out that those consequences are not quite what you'd think, and even though increased market power represents a distinct failure in the sense of distorting allocations in the economy, it explained little of the growth slowdown itself.

MEASURING MARKET POWER

Before getting into the connections of market power and the growth slowdown, let's start by establishing that market power grew in the first place. The most straightforward way to do this is to look at the share of GDP that was paid out as economic profits over time. But looking at those economic profits is not easy, because they are not the same thing as reported accounting profits, and so we have to back them out from the data.