The Peopling of Macroeconomics:

Microeconomics of Aggregate Consumer Expenditures*

BY SATYAJIT CHATTERJEE

Since the 1950s economists have been building a theory of aggregate consumer spending, seeking to understand how individual households choose to spend and how their choices change when interest rates, the unemployment rate, and other economic indicators change. Before that time, economists looked for "economic laws" that would explain the connection between one set of economic aggregates and another, without considering the decisions of individual households. Although the process of connecting macroeconomic aggregates to individuals' behavior is far from complete, predictions of aggregate consumer spending are now rooted in predictions of individual behavior. In this article, Satyajit Chatterjee takes readers through a brief historical survey from the early work on the consumption function to the theory of aggregate consumer spending in modern macroeconomic models.

Consumer spending is the largest single expenditure category in the final demand for goods and services, accounting for more than two-thirds of gross domestic product (GDP). A clear understanding of the underpinnings of consumer spending is a valuable asset for central bankers and policymakers. Since the 1950s, macroeconomists have been engaged in building a theory of aggregate consumer spending from the bottom up.[1] In this approach, macroeconomists first seek to understand how individual households choose to spend and how their choices change when interest rates, the unemployment rate, and other indicators of overall economic activity change. The relationship between aggregate consumer spending and indicators of economic activity is then obtained by aggregating the predicted changes in the spending choices of individual households with respect to changes in indicators of overall economic activity.

It was not always so. In the early years of macroeconomics, scholars looked for enduring empirical relationships ("economic laws") that connected one set of macroeconomic aggregates to another without explicit reference to the individual decisions that would make sense of such connections. This was because economists hadn't fully worked out how a household acting rationally in the face of uncertainty would behave over time — the sort of knowledge needed to meaningfully connect macroeconomic aggregates to the millions of individual choices that make up those aggregates. But as economists began to acquire this knowledge, the process of connecting macroeconomics to individuals' behavior started in the 1950s and gathered steam in the 1970s and 1980s. Although the process of integration is far from complete, predictions of aggregate consumer spending are now rooted in predictions of individual behavior.

The attempt to predict aggregate consumer spending by first predicting what individual households would do is

Satyajit Chatterjee is a senior economic advisor and economist in the Research Department of the Philadelphia Fed. This article is available free of charge at www.philadelphiafed.org/research-and-data/publications/.

*The views expressed here are those of the author and do not necessarily represent the views of the Federal Reserve Bank of Philadelphia or the Federal Reserve System.

[1] Aggregate consumer spending is total consumer spending in the economy.

what I mean by "the peopling of macroeconomics." The aim of this article is to give an account of this now half-century-long intellectual endeavor. It is meant to be a (quick!) historical survey that takes the reader from the early work on the consumption function to the theory of aggregate consumer spending in modern macroeconomic models.

GENESIS OF THE CONSUMPTION FUNCTION

The origin of macroeconomics as a distinct sub-field of economics is often traced to John Maynard Keynes's *General Theory of Employment, Interest, and Money*. Published in 1936, the book sought to explain the reasons for the economic depression that gripped the industrialized world after 1929. In the course of doing so, Keynes introduced a theoretical construct he called the consumption function. According to Keynes, the consumption function was the causal relationship between annual aggregate disposable (or after-tax) income and annual aggregate consumer spending.

Keynes asserted that this relationship looked like the brown line shown in Figure 1. Aggregate consumer spending was directly and linearly related to aggregate disposable income. The point at which the brown line crosses the black line gives the income level at which consumer spending is equal to income. To the left of this point, spending exceeds income, and to the right of this point, spending is less than income. Importantly, the relationship between income and spending was a nonproportional one, with higher incomes associated with a smaller ratio of spending to income. To see this, consider the points marked X and Y on the brown line. At point X income is $40,000 and spending is $34,000; at point Y, income is $80,000 and spending is $58,000. Thus, a doubling of income leads to less than a doubling of spending, which means that the ratio of spending to income declines as incomes rise.

Because the consumption function was central to Keynes's analysis, the construct attracted a great deal of attention and soon became the focus of controversy. The problem was that Keynes did not explain how the consumption function could arise from the choices of individual households acting rationally. Instead, he defended his construct as a "psychological law" that accorded well with common sense. In Keynes's favor, the construct seemed to accord with some facts as well: Household-level incomes and expenditures were (roughly) related as shown in Figure 1, with higher income households spending more than lower income households but spending proportionately less of their income than lower income households.

> According to Keynes, the consumption function was the causal relationship between annual aggregate disposable (or after-tax) income and annual aggregate consumer spending.

But the household-level evidence was not definitive because Keynes's consumption function was supposed to hold for *aggregate* consumer spending and *aggregate* disposable income measured at *different points in time*. The issue remained unsettled because data on aggregate consumer spending and aggregate income for different years were not readily available. When the data were eventually assembled, they showed a relationship like the brown line shown in Figure 2. Over a long period of time, the relationship between consumer spending and income was proportional. As illustrated by the points X and Y, a doubling of (per capita) income from $40,000 to $80,000 leads to a doubling of (per capita) consumer spending from $32,000 to $64,000 — something that is not true of the consumption function in Figure 1. Although one might be tempted to gloss over this difference, the difference was important: Keynes's theory assumed that the consumption function looked like the one in Figure 1, not like the one in Figure 2.

The puzzling difference between consumption-income relationships "across households" (*cross-section*) and "across time" (*time-series*) became the focus of macroeconomic research in the 1940s and 1950s. By that time, many economists had accepted Keynes's *General Theory* as being essentially correct, and it became a matter of some urgency to understand why these relationships differed and how both could be true at the same time. Progress came in the form of two studies that pretty much set the stage for research on the aggregate consumption function for the next 30 years. One was by economist Franco Modigliani and the other by economist Milton Friedman. Both contributions earned their progenitors Nobel prizes: Friedman in 1976 and Modigliani in 1985.

RATIONAL CHOICE: AN ENGINE FOR PREDICTION

Both Friedman and Modigliani focused on understanding the relationship between spending and income at the household level, and both sought

FIGURE 1

Keynesian Consumption Function

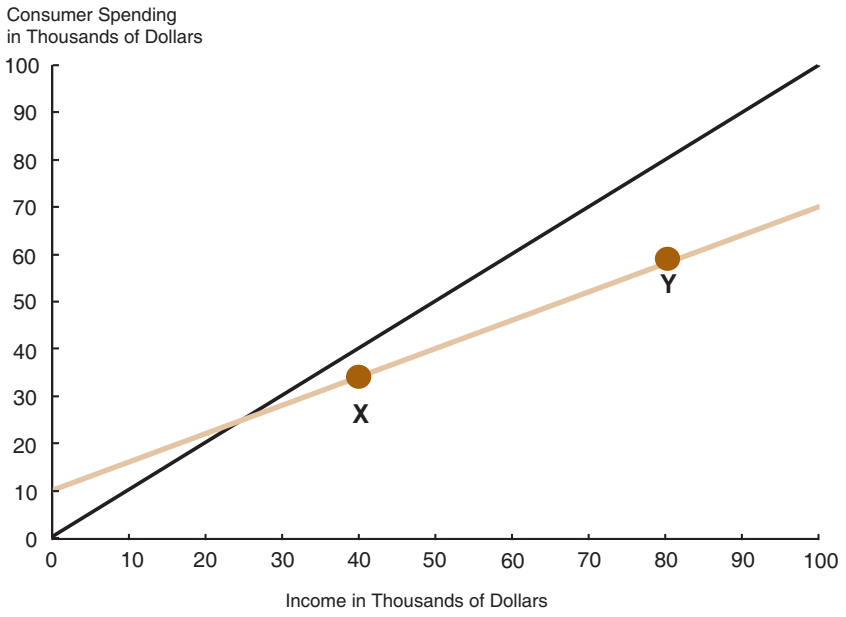
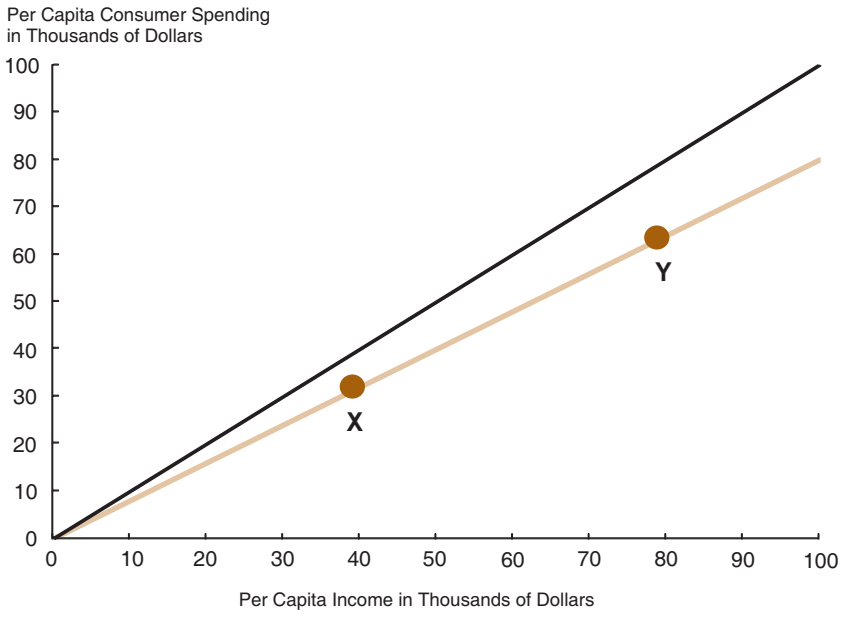


FIGURE 2

Relationship Between Spending and Income Over Time



to achieve this using the model of rational choice. Rational choice — the dominant paradigm for thinking about behavior in economics — posits that people make decisions to maximize their well-being subject to the limitations imposed on them by the finiteness of resources. For instance, the theory of rational choice prescribes how a family should allocate its finite income among competing uses in order to generate the maximum possible well-being for the family.

Modigliani's Work. Modigliani and his student Richard Brumberg began by studying a very simple individual choice problem. They imagined a young adult starting out on his working life at age 20 (say) knowing (with perfect certainty) that he would live up to a given age (say, 90 years), knowing how many of those years he would work (say, 40), and knowing how much he would earn in each period of his working life (say, $35,000 each year). Modigliani and Brumberg assumed that the young adult obtained the same physical and psychological benefit (or utility, as economists call it) from any given amount of spending in any given year. They also assumed that, as is customary in economics, the benefit obtained by the adult from an additional dollar of spending declines with the amount already spent that year: That is, the first dollar spent in any year gives more benefit than the second dollar spent in that year and the second dollar spent in that year gives more benefit than the third dollar spent in that year and so on. Finally, they assumed that the adult could borrow or save at a bank at a zero interest rate.

The question they asked was: What is this individual's best lifetime spending plan? The answer is that the individual should spend his average *lifetime* income of $20,000 each year, where $20,000 is the sum of his in-

come over his working years ($35,000 multiplied by 40) divided by the number of years he will live (70, which is 90 less 20). Because the individual gets the same benefit from spending in each year of his life and because every dollar spent gives less benefit than the previous dollar spent that year, it is best for the individual to spend the *same* amount every year.[2] And if he is to spend the same amount every year and live within his means, he must spend his average lifetime income each year.

Even though the example was highly unrealistic, it served to show that Keynes's consumption function (or "psychological law") had no obvious basis in rational choice. If we could observe this hypothetical individual over time, we would see his income change from $35,000 to zero when he retires and yet we would see his spending stay unchanged at $20,000 a year. Contrary to Keynes's assertion, an increase or decrease in income need not be accompanied by an increase or decrease in spending. In a rational choice context, current spending need not respond to a change in current income if that change is fully anticipated in a previous period.

While these findings raised doubts about Keynes's "psychological law," they did not resolve the issue of the descriptive realism of the "law." Perhaps the "law" was a better description of reality than rational choice. To be truly convincing, proponents of rational choice had to show that their theory explained the facts better than other alternatives. To prove their point, both Friedman and Modigliani concentrated on reconciling the differences between cross-section and time-series consumption-income relationships.

Contrary to Keynes's assertion, an increase or decrease in income need not be accompanied by an increase or decrease in spending.

Modigliani and Brumberg's simple model is consistent with the differences seen in the data between time-series and cross-section consumption-income relationships. In their model, economic growth causes everyone's average lifetime income to grow over time. Since everyone spends their average lifetime income, economic growth also causes aggregate spending to grow at the same rate as average lifetime income. Therefore, spending and income grow in proportion to each other. In contrast, the relationship between income and spending across people alive at any point in time will be necessarily nonproportional because even people without any income (retirees) spend a positive amount (for more details on this point, see *Reconciling Secular and Cross-Section Consumption Functions*).

Of course, this is a simple example, and one might wonder whether the rational choice paradigm would predict these relationships in more realistic situations. The answer to this question is a resounding yes, and the person most responsible for showing why was Milton Friedman.

FRIEDMAN'S PERMANENT INCOME HYPOTHESIS

In 1957 Friedman published a monograph titled *A Theory of the Consumption Function*. As an enduring example of the interplay between economic theory and facts, the treatise has few equals.[3] Friedman distinguished between a household's permanent income and its actual income and — with the help of rational choice theory and empirical facts — argued that a household tends to spend its permanent income.[4]

Friedman defined permanent income as the amount a household could spend and still maintain its wealth. To understand what this definition means, it is helpful to think of some simple examples. First, imagine a household, such as a new retiree, that in terms of resources has only financial wealth. Suppose that a household has a million dollars in the bank, and the interest rate available at the bank is 5 percent. Then, this household's annual *perma-*

[2] To see why, imagine that the individual plans to spend $50,000 in 2008 and only $40,000 in 2009. Because the additional benefit from each dollar spent is declining with the total amount spent, the benefit obtained from spending the 40,001st dollar in 2008 is more than the benefit obtained from spending the 50,000th dollar in 2008. Since the benefit obtained from spending the 40,001st dollar in 2008 is the same as the benefit obtained from spending the 40,001st dollar in 2009, the individual can increase his total benefit by reducing his expenditures by $1 in 2008 and increasing it by $1 in 2009: The loss in benefit in 2008 will be more than compensated by the gain in benefit in 2009. This sort of logic can be applied repeatedly to conclude that the best the individual can do is spend the same amount each year.

[3] To quote Friedman's Nobel citation: "From a purely scientific viewpoint, one of Friedman's most important contributions is his reshaping of consumption theory with the help of the hypotheses about 'the permanent income', in place of current annual income, as a decisive factor in determining total consumption expenditure. Here an extremely fruitful distinction is made between households' temporary income and more permanent income; Friedman shows that a substantially larger part of the former income is saved than of the latter. Friedman has carefully tested this theory on comprehensive statistical material and gained interesting results. Friedman's version of the consumption function has had a lasting effect both on theory and on empirical research."

[4] This is simplifying matters somewhat. Friedman's permanent income hypothesis is the assertion that a household's planned level of spending will be some proportion of its permanent income, where the proportion could fluctuate around unity over time.

nent income is $50,000 — the amount the household would earn in interest and therefore could spend without reducing or augmenting its (financial) wealth.

The example above imagined a household, such as a new retiree, with only financial wealth. What about a young household that has no financial wealth but expects to earn income for many years into the future? Suppose a household expects to earn $40,000 in each of the next 20 years and $60,000 in each of the following 20 years (after which it retires). Suppose it can borrow from the bank against this income stream at an interest rate of 5 percent. Then it is as if this household has financial wealth of (roughly) $820,000 in the bank today — which is the discounted value of the household's stream of future earnings.[5] Then the same logic as above applies, and the household can spend about $41,000 annually — which is (roughly) the annual interest earned on $820,000 — and still maintain its wealth.

Why would a household wish to spend its permanent income? Note that when thinking about how much a household should spend from one month to the next, it is fine to imagine that a household's circumstances are similar from one month to the next. Thus, all else being the same, the household should spend the same amount each month. Second, note that even though a household will exist for a finite length of time, for practical purposes it is fine to imagine that there is no natural end to the household's planning horizon. This may be because the end is really far away or because the household cares about its descendants and its descendants' descendants and so on, so that there is literally no end to its planning horizon. Thus, a household's decision problem is to use a finite amount of wealth to provide for spending over infinitely many future months. The only way this household can spend the same amount each period forever is to spend the constant interest earned on its financial wealth each period, that is, spend its permanent income.

A household's decision problem is to use a finite amount of wealth to provide for spending over infinitely many future months.

One can see how Friedman's permanent income theory could account for the proportional spending-income relationship over time and the nonproportional relationship across households at a point in time. If people's perceptions of their permanent incomes rise with the general rise in living standards, everyone's spending will rise in proportion to the rise in living standards. But if we look at households at a point in time, there will be some households whose income is temporarily above their permanent income, and those households will save most of the additional income; there will also be households whose income is temporarily below their permanent income, and they will draw down their savings to maintain their consumption. Therefore, the relationship between income and spending across households at a point in time will naturally tend to be nonproportional.

Friedman was aware that this approach to consumer spending needed to be amended when uncertainty about future earnings is taken into account. Because a household cannot perfectly forecast its future earnings and because banks do not lend against the promise of uncertain future earnings, there is no way for a household to actually convert its future income stream into an equivalent amount of financial wealth. Nonetheless, Friedman maintained that there must be some notion of permanent income to which household spending is adapted. The level of this permanent income will be household specific and will depend on such things as the household's expected earnings and the household's perception of future earnings risk as well as the household's stock of financial assets.

Although uncertainty about future earnings played a key role in Friedman's theory, the implications of such uncertainty for rational choice were only dimly understood at the time. Friedman did not provide a rigorous foundation for his ideas. The result was that while macroeconomists quickly accepted the distinction between actual and permanent income, they ignored Friedman's assertion that permanent income was something not directly observable. Instead, they took permanent income to mean the annual interest earned on the sum of financial and human wealth, where human wealth was calculated as the present discounted value of current and future *expected* earnings. A key reason behind the adoption of this particular definition was the discovery — made in the 1960s — that under certain conditions the theory of rational choice implied that households should set current

[5] Discounted (or present) value refers to an amount of money today that will become a given amount at a stated point in the future, depending on the interest rate. For example, if the interest rate is 10 percent, $100 today will be worth $110 one year from now. So the present value of $110 one year from now (when the interest rate is 10 percent) is $100.

Reconciling Secular and Cross-Section Consumption Functions

Suppose that each individual lives for two years. An individual works for the first year of his life and enjoys retirement in his second and final year. Each year, one one-year-old is "born" and one two-year-old "dies," so that the total population is always constant at two. There is growth in incomes over time: Every year, newborns earn 20 percent more than the previous year's newborns.

TABLE

Income and Spending in a World of Overlapping Generations

	Year 0		Year 1		Year 2		Year 3	
	Inc.	Spend.	Inc.	Spend.	Inc.	Spend.	Inc.	Spend.
Gen 0	100	50	0	50				
Gen 1			120	60	0	60		
Gen 2					144	72	0	72
Average	…	…	60	55	72	66	…	…

The table records the relevant data for this hypothetical economy. In the table, columns represent either income (Inc.) or spending (Spend.) for a particular year. The generation born in year 0 is denoted Gen 0, the generation born in year 1 is denoted Gen 1, and so on. Thus, under the income column for year 0, there is an entry for 100 in the row representing Gen 0 because that is what the person born in year 0 earns in that year. Moving across the same row, the entry under the spending column in year 0 is 50 because that is what Gen 0 spends in year 0 (the rest of his or her earnings are saved). Continuing to move across, the corresponding entries for year 1 are 0 and 50, respectively, because Gen 0 retires in year 1 and earns nothing but spends 50 in year 1 (this spending is financed by savings accumulated in year 0). Finally, there are no entries for Gen 0 for years 2 and 3 (and beyond) because Gen 0 is not alive in those years. Moving down to Gen 1, there are no entries for year 0 or year 3, since Gen 1 is not alive in those years. For year 1, the entry under the income column is 120 because Gen 1 earns 20 percent more than Gen 0. Gen 1 spends 60 in year 1, and this is recorded under the spending column for year 1. For year 2 the corresponding entries for income and spending are 0 and 60, respectively. The situation is similar for Gen 2. Gen 2 earns 20 percent more than Gen 1 in year 2 and spends half of his earnings in year 2 and the remaining half in year 3. Naturally, there are no entries for Gen 2 for years 0 and 1.

We can use the snapshots of the overlapping generation world displayed in the table to reconcile the shapes of the consumption functions across households (cross-section) and across time (time series). First, let's look at how aggregate per capita income and spending evolve in this economy. The bottom row of the table reports the average income and spending in each of the years for which these averages can be computed from the information reported in the preceding rows.

Let's look at year 1. Aggregate per capita income in year 1 is simply income averaged over the two individuals alive in year 1. The two individuals alive in year 1

FIGURE A

Per Capita Spending and Income in the Overlapping Generations Example

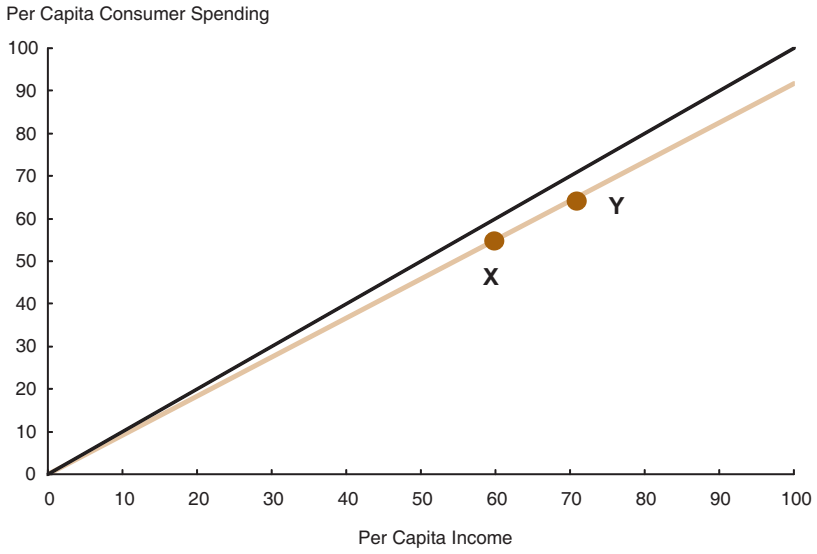
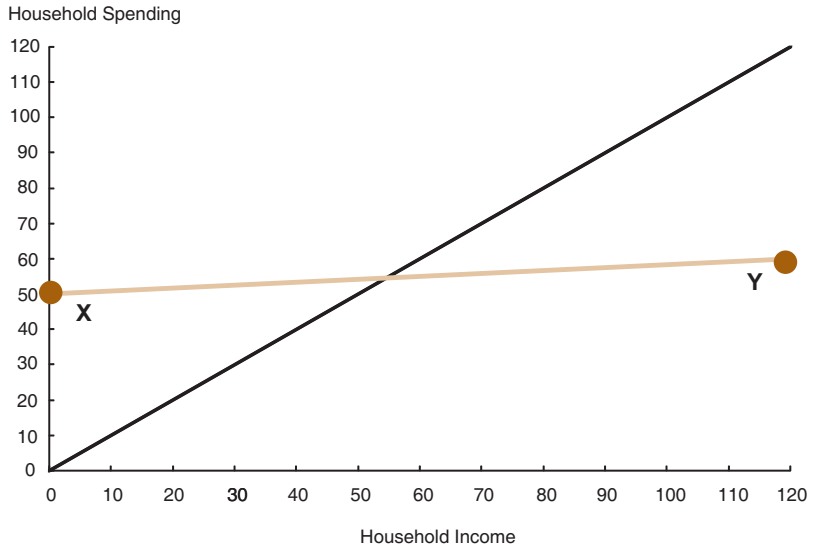


FIGURE B

Spending and Income Across Households in the Overlapping Generations Example



are Gen 0 and Gen 1. Gen 0 has no earnings in year 1 (because he or she is retired) and Gen 1 earns 120 units. So, the average income in year 1 is 60 units (the sum of 0 and 120 divided by 2). Similarly, the aggregate per capita spending in year 1 is 55 units (the sum of 50 and 60 divided by 2). Thus, aggregate per capita consumer spending in year 1 is 11/12 of aggregate income in year 1.

In year 2, the two individuals alive are Gen 1 and Gen 2. Aggregate per capita income is 72 units (the sum of 0 and 144 divided by 2), and aggregate per capita spending is 66 units (the sum of 60 and 72 divided by 2). Once again, aggregate per capita consumer spending is 11/12 of aggregate per capita income. Figure A plots aggregate per capita spending and income at successive points in time for this economy. As is evident, income and spending grow in proportion to each other over time exactly as found in the data.

Next, let's look at the cross-section consumption-income relationship in this economy. Let's pick year 1. Gen 0 has no income and spends 50 units, and Gen 1 earns 120 units and spends 60 units. Therefore, the cross-section consumption-income relationship for year 1 looks like the one in Figure B. This relationship is clearly not proportional and, in fact, resembles the consumption function in Figure 1 in the text. If we were to pick a different year, say, 2, we would get a similar nonproportional relationship except that it would be shifted upward because of income growth.

spending equal to permanent income calculated in this way. This resulted in this particular definition becoming commonly used, and eventually, the very idea of permanent income became associated with this particular definition.

But this interpretation of permanent income turned out to be inconsistent with the evidence. As more extensive aggregate and household-level data became available for macroeconomists to analyze, it was found that consumer spending responded too much to transitory deviations in income from permanent income defined in this way to be consistent with the theory's predictions.[6]

UNCERTAINTY, BUFFER-STOCK SAVINGS, AND SPENDING DYNAMICS

While macroeconomists were busy testing the permanent income theory against aggregate and household data and finding it wanting, others were concentrating on working out the implications of rational choice for decision-making over time when the future could not be perfectly forecast. The big hurdle here was that it was not easy to divine the full implications of rational choice theory because the theory's predictions could not be reduced to a simple formula. Consequently, it was not easy to figure out if some version of rational choice theory could explain the data on household spending and income better than the permanent income theory.

Two key developments eventually allowed progress to be made. The first development was something not intrinsically connected to economics. It was the increasing availability of (and access to) high-speed computers on university campuses and the concurrent rapid development and standardization of computer languages designed to express and solve difficult numerical problems. Along with a deeper understanding of the nature of rational choice over time, the rapid improvement in the hardware and software for numerical computations permitted macroeconomists to pose, solve, and simulate rational choice problems on the computer.

The other, more important development was connected with the progress of economics as a discipline. To solve rational choice problems on the computer, one must specify the problem in exact numerical form. For instance, it is no longer sufficient to assert (as Modigliani and Brumberg did) that the benefit from an additional dollar of spending declines with the amount already spent; it is necessary to specify how much it declines at any given level of spending. In other words, the computer needs to know the exact numerical relationship between the benefit from an additional dollar of spending and the level of spending.

It took macroeconomists decades to gather this kind of knowledge. The process was helped by the fact that the rational choice paradigm had become the *lingua franca* of economics — the common language economists use to make sense of behavior in diverse branches of economics. For instance, one source from which macroeconomists learned of the numerical relationship between the benefit from an additional dollar of spending and the level of spending was researchers trying to understand how fluctuations in expected rates of return on financial assets affected the growth rate of consumer spending.

While macroeconomists were busy testing the permanent income theory against aggregate and household data and finding it wanting, others were concentrating on working out the implications of rational choice for decision-making over time when the future could not be perfectly forecast.

Through this process, it became possible for macroeconomists to explore the implications of rational choice for consumer spending using computer simulations. Christopher Carroll and Angus Deaton were among the pioneers of this research effort. Their simulations revealed that households that start without any financial assets initially consume less than their earnings in order to accumulate a *buffer* stock of savings.[7] They do so because savings can protect the household from temporary shortfalls in earnings. Since earnings in any period (a year or a month) are uncer-

[6] See the article by Robert Hall and Frederic Mishkin.

[7] The discussion in the rest of this section draws on Christopher Carroll's article.

tain, there is value to having such a buffer stock of savings. Then, once the household accumulates its target buffer stock of savings, it acts to maintain that stock over time. Unexpected increases in income are initially saved but then gradually spent to bring the stock of savings down to its target level. Similarly, an unexpected decline in earnings is initially met by a reduction in the stock of savings (as the household tries to maintain spending), but then the resulting deficit in its stock of savings is gradually made up over time.[8] Finally, starting at about age 50, behavior undergoes a significant change: While still working, households rein in their spending and begin to accumulate additional savings to provide for their retirement.

The behavior revealed by these computer simulations has a simplicity to it that gives it a ring of truth. But what makes these predictions compelling for macroeconomists is that the simulations also explain why spending's response to transitory fluctuations in earnings can be larger than that predicted by the permanent income theory. When households are working toward accumulating their target level of buffer-stock savings, their spending is depressed. The simulations reveal that in these circumstances a household that receives an unexpected transitory increase in income has an incentive to boost spending from its depressed level. This happens because the extra income is used to augment the household's savings, and therefore, the household gets closer to (or achieves) its target level of buffer-stock savings. Consequently, the incentive to curtail spending in order to get to the target level of buffer-stock savings is attenuated or eliminated, and spending responds strongly to a transitory increase in income. This effect is absent in the permanent income theory because households do not curtail their spending in order to accumulate a buffer stock of savings.

[8] It is worth noting that macroeconomists were aware that rational choice theory was consistent with households' accumulating assets in order to meet a potential shortfall in earnings in the future. What the simulations revealed — and this came as a surprise — was the centrality of precautionary or buffer-stock savings in the household's spending decisions.

When households are working toward accumulating their target level of buffer-stock savings, their spending is depressed.

Interestingly, the simulations also reveal that the predictions of the permanent income theory continue to be relevant once a household reaches its target level of savings. As the buffer-stock of savings is approached, households act more like future uncertainty does not matter – just as the permanent income theory had assumed. Of course, the household behaves this way because it has accumulated a buffer stock of savings to counter the risk of lost earnings. That being said, it is important to note that because households are continually buffeted by shocks to earnings and are therefore accumulating or drawing down financial assets, the fraction that behaves according to the permanent income theory in the simulations is a minority.

CYCLICAL IMPLICATIONS OF BUFFER-STOCK SAVINGS MODELS

Macroeconomists and policymakers are interested in what Deaton's and Carroll's spending models have to say about aggregate consumer spending and savings as well as the movement of these aggregates over the course of business cycles. As one might suspect, the only way to get answers to these questions is by computer simulations. But the simulations are no longer about the behavior of a typical household but the aggregate behavior of an entire *ensemble* of households, an ensemble whose summed behavior has measurable effects on the cyclical behavior of market prices and interest rates. Since the cyclical behavior of market prices and interest rates, in turn, affects the behavior of each household in the ensemble, the challenge for the simulation is to properly account for the feedback from behavior to market prices and back to behavior.

The "feedback" problem prevented macroeconomists from analyzing the business-cycle implications of buffer-stock savings behavior until, in an important paper, Per Krusell and Anthony Smith showed how the problem could be solved. They developed a procedure for reliably compressing the amount of information required by the computer to keep track of feedback effects. With this innovation, macroeconomists are now able to simulate the behavior implied by rational choice of a large ensemble of interacting households living through expansions and recessions.

The simulations reveal that cycli-

cal fluctuations in aggregate consumer spending and aggregate income are more tightly linked than the permanent income theory implies. This makes intuitive sense: The tighter link is a consequence of the fact that those households whose spending is depressed because they are in the process of accumulating their target level of buffer-stock savings will increase their spending more when income is temporarily high (as it is in an expansion).

CONCLUSION

Macroeconomics studies the structure and performance of an economy as a whole. Although the founding documents of economics have a decidedly macroeconomic focus — Adam Smith wrote about the wealth of nations, after all — the development of economics as a modern discipline has been a long and arduous effort to understand and predict the behavior of individual decision-making units, such as households and business firms.

For historical reasons, macroeconomics began life with a rather tenuous connection to the principles of rational choice, in part because John Maynard Keynes explicitly rejected rational choice – and its correlate of competitive markets – as a framework unsuitable for explaining the Great Depression. But it was also because of the broad scope and general complexity of the subject matter; it is a field that invites theorizing at the macro rather than at the micro level.

But fortunately for the development of macroeconomics, there was one very important point of contact between macro- and microeconomics, namely, the consumption function. To make sense of this function, macroeconomists had to think seriously about individual behavior. And so began the "peopling of macroeconomics." The process has gone on now for more than 50 years, and, to quote Angus Deaton, it has "generated some of the best science in economics." This article has endeavored to give a glimpse of this fascinating and ongoing intellectual journey.

REFERENCES

Carroll, C. "Buffer Stock Savings and the Life-Cycle/PIH.," *Quarterly Journal of Economics*, 112 (1997), pp. 1-56.

Deaton, A. "Savings and Liquidity Constraints," *Econometrica*, 59 (1991) pp. 1221-48.

Friedman, Milton. *A Theory of the Consumption Function*. Princeton: Princeton University Press, 1957.

Hall, R.E., and F.S. Mishkin. "The Sensitivity of Consumption to Transitory Income: Estimates from Panel Data on Households," *Econometrica*, 50 (1982), pp. 461-81.

Krusell, Per, and Anthony Smith. "Income and Wealth Heterogeneity and the Macroeconomy," *Journal of Political Economy*, 106:5 (1998), pp. 867-97.

Modigliani F., and R. Brumberg. "Utility Analysis and the Consumption Function: An Interpretation of Cross-section Data," in Kenneth K. Kurihara, ed., *Post-Keynesian Economics*. New Brunswick, N.J.: Rutgers University Press, 1954, pp. 388-436.

Accounting for Cross-Country Differences In Income Per Capita*

BY AUBHIK KHAN

Living standards, as measured by average income per person, vary widely across countries. Differences in income result in large disparities in spending on goods and services by people living in different economies. What makes some countries rich and others poor? Furthermore, what determines income per person in a country, and why are these factors unevenly allocated across the world? In this article, Aubhik Khan outlines a framework for growth accounting to account for cross-country differences in income. The current consensus is that differences in per capita income across countries don't arise primarily from differences in the quantities of capital or labor, but rather from differences in the efficiency with which these factors are used.

Living standards, as captured by average income per person, vary dramatically across countries. These differences in income result in large disparities in spending on goods and services by people living in different economies. The typical person in a poor country has not only less consumption of food and housing but also less education and poorer health, when compared with a typical person living in a rich country. There are also sharp differences in life expectancy and infant mortality between rich and poor countries, both falling with income per capita.

In an effort to illustrate the magnitude of these differences in income, let's examine real gross domestic product (GDP) per capita using the cross-country data available from the Penn World Tables.[1] For 2004, the most recent year for which this measure exists, there are data on per capita GDP for 82 countries. Aside from Luxembourg, which is anomalous, the richest fifth in this group have an average income per capita of $32,142.[2] The poorest fifth have an average income per capita of only $1,422. Thus, the 16 poorest countries for which we have data for 2004 have an average income that is 23 times less than that of the 16 richest countries. This means that the typical person living in these poorer countries must survive on $4 each day. In the absence of government subsidies, it is difficult to imagine how an individual could buy enough food and shelter in the U.S. to survive with this income.

What makes some countries relatively rich while others are unimaginably poor? More generally, what are the determinants of income per person in an economy, and why are these inputs allocated so unevenly across the world? Why are some countries always at the bottom of the tables, while others rapidly close the gap between themselves and richer nations? We are compelled to ask

Aubhik Khan is an associate professor of economics at Ohio State University. When he wrote this article he was a senior economic advisor and economist in the Philadelphia Fed's Research Department. This article is available free of charge at www.philadelphiafed.org/research-and-data/publications/.

*The views expressed here are those of the author and do not necessarily represent the views of the Federal Reserve Bank of Philadelphia or the Federal Reserve System.

[1] The Penn World Tables, prepared by Alan Heston, Robert Summers, and Bettina Aten, facilitate cross-country comparisons by calculating real GDP per capita for a large set of countries using a common set of international prices. It is widely used for cross-country comparisons because it assigns the same value to any particular commodity or service regardless of country.

[2] Luxembourg is anomalous not only because of its size but also because its income per capita, $54,285, is far beyond that of the rest of the rich world. The next richest country, the United States, has an income per capita of $39,535.

such questions because their answers might give policymakers a chance to implement a dramatic improvement in living standards in poorer countries. Nobel laureate Robert E. Lucas writes: "The consequences for human welfare involved in questions like these are simply staggering: Once one starts to think about them, it is hard to think about anything else."

Economists have studied sources of cross-country differences in income for more than 200 years. In the 1950s, Nobel laureate Robert Solow developed a framework for growth accounting that has been used extensively by economists to account for cross-country differences in income. Researchers in this field have achieved a remarkable degree of consensus that differences in per capita income across countries don't arise primarily because of differences in the quantities of capital or labor but rather because of differences in the efficiency with which these factors are combined. Further research on the underlying sources of these differences has provided further insights.

ACCOUNTING FOR CROSS-COUNTRY DIFFERENCES IN INCOME PER CAPITA

Accounting for cross-country differences in income is a daunting task. Why is one country richer or poorer than another? One could think of a host of reasons involving differences in skills; technologies; economic policies; natural endowments, including land, climate, and the frequency of natural disasters; political stability and human rights; the role of women in the workforce; and many other phenomena.

Whether studying the reasons for changes in Great Britain's income over the course of the Industrial Revolution or why Bangladesh is poorer than Thailand, economists begin by studying production in each country. The total value of all goods and services produced in the nation — real GDP — can be attributed to one of three sources: capital, labor, and total factor productivity. The manner in which differences in the levels of these factors translate into differences in real GDP is determined through the aggregate production function.

Whether studying the reasons for changes in Great Britain's income over the course of the Industrial Revolution or why Bangladesh is poorer than Thailand, economists begin by studying production in each country.

AGGREGATE PRODUCTION FUNCTION

Before describing the production function, let's review the factors of production listed above. At the simplest level of aggregation, capital and labor are arguably always present in the production of any commodity — whether restaurant meals, economics lectures, or other goods and services.

Capital. Capital is the sum of all different types of equipment and structures used in production. Examples of equipment include both ploughs and tractors and both motorcycles and buses. This suggests the first problem in growth accounting, one that affects all of macroeconomics: How do you add up different goods to arrive at a total stock? If we want a single measure of all of the capital in the economy, how many ploughs make a bus? We need a way to assess the value of each commodity. Economists often use market prices as a measure of value. Thus, if a plough costs $1000 and a tractor costs $15,000, an economist would assume that the value of a tractor, as capital, is 15 times the value of a plough. A hypothetical economy that had only ploughs and tractors, 10 of the first and two of the second, would have a total capital stock of $40,000.

Simple aggregation as described above cannot be directly applied to the measurement of capital because we don't count the quantities of different types of capital existing in an economy. We don't know how many ploughs there are in Great Britain because there is no direct measurement of existing stocks. In contrast, there is direct measurement of flows. We count the output of every firm, and thus we have a good estimate of how many new ploughs are made each year. Thus, while we lack data on the stocks of capital, we do have data on investment in these stocks.

Economists infer a measure of capital stock through the aggregate flow of investment using what is known as the *perpetual inventory method.* In its simplest application, this assumes that all capital goods lose a constant fraction of their value as they deteriorate through use. Known as physical depreciation, this notion captures both breakdowns and obsolescence, not only of machinery but of all forms of capital. The existence of depreciation implies that there must be gross investment to simply maintain the existing capital

stock because some of it is lost each year. A conventional estimate of the average depreciation rate for the United States is around 6 percent.

The capital stock this year is calculated as the sum of the nondepreciated fraction of capital from last year and current gross investment. This method relies on an initial guess for capital, but depreciation reduces the importance of this guess over time. The perpetual inventory method determines the total capital stock existing today as the weighted sum of all past investments, with the weight on past investment declining over time because of depreciation.

Aside from physical capital, economists have also tried to address cross-country differences in intangible capital.[3] Examples include spending on research and development, training employees, creating new businesses and other forms of organizational capital, and the accumulated experience and know-how of productive organizations. Most of these investments in intangible capital are not counted in national income and product accounts. This omission understates the importance of broad capital in production.

Labor. Labor is as diverse as capital. In most studies of cross-country income differences, labor input is measured as the total stock of human capital. Human capital is simply the quality-adjusted stock of workers, just as physical capital is the stock of equipment and structures used in production. The stock of human capital in an economy divided by the number of workers gives an average measure of the skill of the workforce.

This leaves open the issue of how to measure the average level of skills. One common approach is to use cross-country data on the average years of schooling provided by Robert Barro and Jongwha Lee in 1993. However, a person's years of schooling are not a measure of his or her skill but a measure of the quantity of time invested in acquiring skills. To convert years of schooling into a level of human capital, the returns to schooling are often used. This approach, developed by Jacob Mincer and described in his 1974 book, assumes that people with higher levels of human capital are paid higher wages because they are more productive in their jobs and, as a result, more valuable to their employer. By examining the increase in real wages arising from an additional year of schooling, economists can use what are known as *Mincerian regressions* to convert the average years of schooling in a country into a measure of average human capital per worker. Typical estimates of the increase in real wages from an additional year of schooling are about 10 percent.

TFP represents the efficiency with which inputs, capital and labor, are used. As such, it is often interpreted as a measure of the effectiveness of the technologies used in an economy.

Output. After the measurement of inputs, we have to address the measurement of output. The production of different goods and services is summed using market prices, whenever possible, as a measure of their value. This is similar to the approach described to aggregate the capital stock. It allows us to arrive at aggregate measures of total production in economies that produce an amazing diversity of goods and services. In what follows below, we will use per capita GDP, the level of goods and services produced per person, interchangeably with income per person.[4]

Total Factor Productivity. There is one last element in the aggregate production function. Economists have found that two countries with identical levels of capital and labor do not produce identical levels of output. More generally, given the stock of capital and labor, the level of output produced by these inputs varies substantially.[5] This variation exists both over time within a country and across countries at a specific time. This phenomenon is described as variation in *total factor productivity* (TFP). TFP represents the efficiency with which inputs, capital and labor, are used. As such, it is often interpreted as a measure of the effectiveness of the technologies used in an economy. Economies with higher TFP are believed to produce using more efficient technologies that provide

[3] The 2002 book by Stephen Parent and Edward Prescott provides a more extensive discussion of the issues involving the measurement of intangible capital. They conclude that differences in intangible capital cannot, by itself, explain much of the cross-country differences in income.

[4] This is not exactly correct because income per person is actually better represented by gross national product, or GNP, rather than GDP. The difference between these two measures of income per person arises when the citizens of a nation have sources of income from production outside their own nation. Furthermore, in the Penn World Tables, the market prices are actually international prices based on a weighted average of prices calculated for each country.

[5] This is also known as the Solow residual, since it was first isolated by Robert Solow.

more goods and services for any given level of capital and labor.

TFP is not directly measured. Instead, its level is determined by dividing GDP by a benchmark level of output, that is, the level of output that would exist if TFP were one.

But how do we know the level of output when TFP is one? This is where the aggregate production function enters the analysis, providing a benchmark measure of output from the factors of production: capital and labor. Many forms of the production function have been used in economics, but growth accounting usually uses the Cobb-Douglas production function. (See *The Cobb-Douglas Production Function*.)

Permanent increases in an economy's TFP are thought of as technological progress. This is because such a change implies that the economy can produce more output using the same stocks of physical and human capital. In other words, the economy is using a new, more productive technology.

In truth, aside from differences in the level of technology, TFP differs across countries for many other reasons. Differences in other factors of production, not directly measured, are just one such reason. Thus, the levels of raw materials and energy used in production are implicitly captured by TFP. If two countries have the same capital and labor, but the first has twice the level of raw materials and energy as the second, then TFP will be higher in the first country than in the second. As we shall learn below, much recent research into growth accounting focuses on the causes of differences in TFP.

CROSS-COUNTRY DIFFERENCES IN INCOME

Economists are primarily interested in explaining differences in income per person, or, more formally, real GDP per capita. It is, of course, no mystery if a country twice the size of another produces twice as much. All else equal, this would arise simply because one country had twice the number of people, and thus twice the workers, of the other. There need be no difference in TFP or capital per worker. The question of why one country contains twice the people compared with another country may still be of interest to social scientists. However, the more limited goal of growth accounting is to explain differences in income per person.

A simple reshuffling of the aggregate production function allows us to attribute production per person to either capital per person, TFP, or the average level of human capital in an economy. In this way, we can use the aggregate production function described above to break down cross-country differences in income and, as a result, to begin to answer the primary

The Cobb-Douglas Production Function

When computing the level of output that will be produced given a stock of capital and level of labor hours, economists often apply a relationship known as the Cobb-Douglas production function. If Y is used to denote output, K is the variable that represents capital, and L stands for labor, the Cobb-Douglas production function is the relationship:

$$Y = AK^{\alpha}L^{1-\alpha}$$

Here α is a coefficient between 0 and 1 that captures the percentage change in output that results from an additional unit of capital. It is also known as capital's share. Similarly, in the above version of the Cobb-Douglas function, labor's share is $1-\alpha$. The sum of shares is then equal to 1, which implies that if we increase both capital and labor by some proportion, output will also rise by that same proportion.

The share term, α, is calculated using data on either the income earned by capital or the income earned by labor. Under the assumption that factors of production are paid competitively, the share of total production that is paid to labor will equal $1-\alpha$.

If there is imperfect competition, and firms have monopoly power, then $1-\alpha$ will exceed the share paid to workers. However, provided we have a measure of firms' markups of price over cost, we can still use labor income data to derive the coefficient, α.

Given the direct measurement of output, Y, the capital stock, K, the stock of human capital, L, and the coefficient α, the level of TFP is given by A. It is the fraction of output that cannot be explained by the stock of capital and labor.

The form of the Cobb-Douglas production function implies that, in competitive markets, the share of income paid to capital and labor will be constant. This is broadly supported by empirical evidence showing that, over long periods of time, there has been little change to the share of income paid to labor and capital.

question of economic development: Why are some countries richer than others?

Differences in TFP Explain Most of the Variation in Income. The tangible wealth of a nation is contained in its physical capital; intangible wealth lies in human capital and in TFP. A reader may have believed that most of the differences in income per person across countries may be explained by differences either in the quantity of physical capital per worker or in the skills of the workforce. However, the startling finding from growth accounting over the past decade is that the majority of cross-country differences in income per person arise through differences in TFP. Most researchers believe that measurable inputs such as physical and human capital explain less than half of the difference in income. Rather, it is the level of technology used that explains the majority of this difference. While the list of researchers who have made important contributions to this debate is lengthy, three influential papers are the 1997 work by Peter Klenow and Andrés Rodriguez-Clare, the 1998 lecture by Edward C. Prescott given at the University of Pennsylvania, and the 1999 study by Robert Hall and Chad Jones. Across these studies TFP is found to explain between 50 and 75 percent of the observed differences in income per capita.

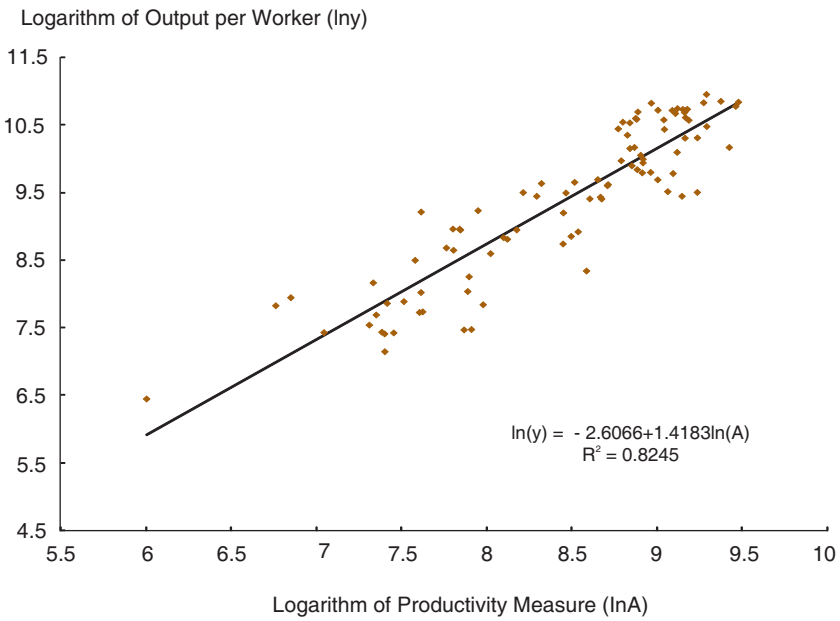
The figure, which is derived using data made available by Francesco Caselli, shows the relationship between TFP and income per capita in 1996. As explained by Hall and Jones, who derived a similar figure using 1988 data, the figure shows that the differences in income across countries is very similar to the corresponding differences in total factor productivity — that fraction of output that cannot be explained by capital and labor. The correlation between output per worker and total factor productivity exceeds 80 percent.

Differences in Capital and Labor Are Less Striking. If differences in technology, as captured by TFP, are the primary determinant of differences in income, physical and human capital are less important explanatory variables. It is certainly true that richer economies have more capital per worker. However, the extent of the cross-country variation in capital per worker is not large enough to explain most of the observed differences in income. To see this, we again use the data set developed by Francesco Caselli for his chapter in the *Handbook of Economic Growth.* Across the 94 countries in his sample, the richest 20 percent had income per capita that was almost 22 times that of the poorest 20 percent. However, after he adjusted for the importance of capital in production, the differences in the ratio of capital to output across these two groups of countries was somewhat less than two-fold.

Still there is more capital used in richer countries, and the underlying reasons for this are an important issue. However, this does not appear to be because savings or investment rates are higher in richer countries. As shown by Chang-Tai Hsieh and Peter Klenow in their 2007 paper, when measured using domestic prices, savings rates do not vary systematically with average income. Instead, it appears that poorer countries are less efficient at producing investment goods relative to goods used for consumption.

It's also true that richer countries have higher levels of skills per worker. However, the Mincerian approach to calculating skills does not lead to cross-country variation in the stocks of human capital, which suggests a much larger role for human capital in explaining income differences than that found for physical capital.

FIGURE



Source: Data set from Francesco Caselli "Accounting for Cross Country Differences in Income," and available at http://personal.lse.ac.uk/casellif/.

Returning again to the data used by Francesco Caselli, the ratio of average human capital in the richest fifth of nations, relative to that in the poorest fifth, was about two, very similar in size to differences in capital.

Adding It All Up. The apparent unimportance of measurable inputs leads to the following conclusion. In general, to explain why one country is poorer than another, you must explain why it has lower TFP. How large are these differences in TFP? The data used by Francesco Caselli suggest that the ratio of TFP between the richest and the poorest 20 percent of countries is more than five-fold. When taken alongside differences in physical and human capital, this explains the difference in overall GDP per capita.

Remember that the ratio of per capita income between the richest and poorest 20 percent of countries is about 20. The Cobb-Douglas production function gives us an accounting identity that breaks this difference down into the product of three other ratios: (i) capital divided by output and adjusted for a term reflecting capital's share of production, (ii) labor, and (iii) TFP. Their values are (i) 1.85, (ii) 2.06, and (iii) 5.36 and their product is $1.85 \times 2.06 \times 5.36 = 20.4$.[6]

Subsequent work re-examining the sources of cross-country income differences has largely confirmed the original findings that TFP explains most of the difference we see. In reaching this consensus, economists have carefully tried to address problems that might arise from errors present in the measurement of inputs and output. These efforts have led to better measures of schooling and more precise calculations of human capital. There have also been corrections for the quality of goods and services produced in rich and poor countries. The implications of different aggregate production functions, other than the conventional Cobb-Douglas, have been studied. Researchers have also corrected for different levels of market versus home production across countries. In poorer economies, more goods and services are produced at home, using time-intensive methods of production, than in the marketplace. Omitting the value of such home production, which is not included in national accounts, exaggerates the income disparity between rich and poor countries. This research is summarized in the survey by Francesco Caselli and in Peter Klenow's 2006 plenary address to the Society for Economic Dynamics.

Subsequent work re-examining the sources of cross-country income differences has largely confirmed the original findings that TFP explains most of the difference we see.

EXPLAINING DIFFERENCES IN TFP

As I've described above, a consensus has developed on the primary importance of cross-country differences in TFP for explaining differences in income per capita. However, the accounting methodology used to arrive at this consensus has presented a problem. Since TFP is inferred as a residual and not directly measured as physical or human capital are, attributing differences in income to differences in TFP does not completely answer the question of why countries differ. All we have really found is that these differences cannot be attributed to measured differences in physical or human capital. They lie somewhere else. Economists have started to examine the causes of differences in TFP across countries.

Looking Behind the Aggregate Production Function. An important early contribution to this research was made by Stephen Parente and Edward C. Prescott, who, in their 1999 paper, described how the adoption of more productive technologies may be hampered because groups of people have vested interests in protecting existing, but less productive, technologies. Following their work, a large body of research has arisen. Some of this work looks inside the production function for the economy. This research seeks to examine how factors affecting the production decisions of individual firms add up to differences in output at the aggregate level. Instead of attempting a full survey of this literature, I mention two recent examples.

One interesting line of research studies how taxes and other distortions, such as employment protection policies, can reduce TFP. For example, in their paper, Diego Restuccia and Richard Rogerson study the effect of taxes and subsidies that favor some firms relative to others. They find that such policies lead to too much capital and labor being used by some plants that benefit

[6] The reason that there is a small difference between the product of these ratios, which is 20.4, and the ratio of per capita GDP between the poorest and richest 20 percent of economies, which is 21.82, is somewhat technical. This discrepancy, a result of something known as Jensen's inequality, arises because the product of the average of the ratios is not equal to the average of the product of the ratios.

from subsidies. By moving capital and labor from productive plants to unproductive plants, such policies can lead to a reduction of between 30 and 50 percent in an economy's TFP. This research provides an example of how TFP is not necessarily determined by technological know-how alone but is also affected by economic policies.

Amartya Lahiri and Kei-Mu Yi also emphasize the role of economic policies in explaining the different economic performance of two Indian states, West Bengal and Maharashtra. Economic development in these two states poses an interesting puzzle. In 1960 West Bengal's GDP per capita exceeded that of Maharashtra, but by 1993 its GDP was barely two-thirds that of Maharashtra. Lahiri and Yi use this case study as a means to get behind the aggregate production function. In their model, there are separate production functions for agriculture, manufacturing, and services. They conclude that West Bengal has fallen behind Maharashtra because TFP in manufacturing and services has grown more slowly. Returning to our language above, there has been less technological progress in West Bengal. Lahiri and Yi suggest that growth in TFP has been lower in West Bengal because labor and industrial regulations have hindered growth in business TFP. In general, policies that stifle innovation or the adoption of new, more efficient technologies slow TFP growth. This, in turn, reduces the growth of income per capita.

Re-examining the Role of Human Capital. Recently, researchers have begun to question the importance of TFP. In two separate papers, Rodolfo Manuelli and Ananth Seshadri and, separately, Andres Erosa, Tatyana Koreshkova, and Diego Restuccia have argued that human capital has not been properly measured in existing studies.

They have two main insights. The first is that human capital investment in a country is not independent of the level of TFP. Second, they argue that human capital investment requires not only years of schooling but also goods and services such as schools and teachers. This, in turn, suggests that years of schooling are an incomplete measure of human capital because the quality of the human capital is neglected.

Households make educational investments based on the returns to education, and these investments involve not only the time spent in school but also real goods devoted to education. This implies that the standard Mincerian approach to inferring the stock of human capital may understate cross-country differences. These authors argue that a different approach to measuring human capital, one where households explicitly undertake schooling decisions that vary across countries in response to the economic environment, leads to much larger differences in quality-adjusted human capital across countries. This, in turn, reduces the direct role of TFP. Indeed, they find that cross-country differences in human capital are the leading source of differences in income. However, it remains true that these differences in human capital are driven by differences in TFP. It's just that the required differences in TFP become far smaller.

CONCLUSION

Economists account for cross-country differences in income per person using the method known as growth accounting. It breaks down real GDP per capita into capital per worker, human capital per worker, and the level of technology, or TFP. TFP is the level of output that remains to be explained after accounting for the role of physical and human capital. Measuring the levels of these inputs across countries, we find that most of the cross-country variation in income per person is attributable to differences in TFP. Poorer economies are poorer not because they have less capital and lower skills per worker but because they use these inputs less efficiently than wealthier economies.

Many things can affect a nation's TFP. For example, economic policies, such as taxes or subsidies, may impede the efficient distribution of capital and labor across firms, which will lower TFP. Alternatively, they may prevent the adoption of the most efficient technologies and thus lower TFP. However, to the extent that the technology is much more readily transferable across countries than physical or human capital, why would one country suffer the loss in output associated with using an inferior technology? If, instead, TFP differs because of policies that hinder the growth of business, why allow such policies to persist when the gains to correcting them are so large?

REFERENCES

Barro, Robert J., and Jongwha Lee. "International Comparisons of Education Attainment," *Journal of Monetary Economics* 32 (1993), pp. 363-94.

Caselli, Francesco. "Accounting for Cross-Country Income Differences," in P. Aghion and S. Durlauf (eds.), *Handbook of Economic Growth*, North Holland (2005).

Erosa, Andres, Tatyana Koreshkova, and Diego Restuccia. "How Important Is Human Capital? A Quantitative Theory Assessment of World Income Inequality," Working Papers tecipa-280, University of Toronto, Department of Economics (2007).

Hall, Robert, and Chad Jones. "Why Do Some Countries Produce So Much More Output per Worker Than Others?" *Quarterly Journal of Economics*, 114 (1999), pp. 83-116.

Heston, Alan, Robert Summers, and Bettina Aten. Penn World Table Version 6.2, Center for International Comparisons of Production, Income, and Prices at the University of Pennsylvania (September 2006); (http://pwt.econ.upenn.edu).

Hsieh, Chang-Tai, and Peter J. Klenow. "Relative Prices and Relative Prosperity," *American Economic Review*, 97 (2007), pp. 562-85.

Klenow, Peter K. "Income Differences Across Countries," plenary address to the Society for Economic Dynamics, July 6, 2006, Vancouver, Canada (available at www.klenow.com/KlenowSED.pdf).

Klenow, Peter, and Andrés Rodriguez-Clare. "The Neoclassical Revival in Growth Economics: Has It Gone Too Far?" in National Bureau of Economic Research *Macroeconomics Annual*, Cambridge, MA: MIT Press (1997).

Lahiri, Amartya, and Kei-Mu Yi. "A Tale of Two States: Maharashtra and West Bengal," Federal Reserve Bank of Philadelphia Working Paper 06-16/R (April 2008); available at http://www.philadelphiafed.org/research-and-data/publications/working-papers/2006/wp06-16.pdf.

Lucas, Robert E., Jr. "On the Mechanics of Economic Development," *Journal of Monetary Economics*, 22 (1988), pp. 3-42. Reprinted as Chapter 1 in *Lectures on Economic Growth*, Cambridge, MA: Harvard University Press (2002).

Manuelli, Rodolfo, and Seshadri, Ananth. "Human Capital and the Wealth of Nations," Working Paper, University of Wisconsin-Madison (April 2007).

Mincer, Jacob. *Schooling, Experience, and Earnings*. New York: Columbia University Press (1974).

Parente, Stephen L., and Edward. C. Prescott. "Monopoly Rights: A Barrier to Riches," *American Economic Review*, 89 (1999), pp. 1216-33.

Parente, Stephen L., and Edward C. Prescott. *Barriers to Riches*. Cambridge, MA: MIT Press (2000).

Prescott, Edward C. "Needed: A Theory of Total Factor Productivity," *International Economic Review*, 39 (1998), pp. 529-49.

Restuccia, Diego, and Richard Rogerson. "Policy Distortions and Aggregate Productivity with Heterogeneous Plants," *Review of Economic Dynamics*, 4 (October 2008), pp. 707-20.

Solow, Robert M. "Technical Change and the Aggregate Production Function," *Review of Economics and Statistics*, 39 (1957), pp. 312-20.

Rethinking the Implications of Monetary Policy:
How a Transactions Role for Money Transforms the Predictions of Our Leading Models*

BY JULIA K. THOMAS

Over the past several decades, economists have devoted ever-growing effort to developing economic models to help us understand how changes in interest rates brought about by monetary policy actions affect the production and provision of goods and services in the economy. Although New Keynesian models have broad appeal in explaining how changes in the money stock can affect business activity, these models generate results that are inconsistent with what we know about how interest rates move with policy-induced changes in the money stock. In this article, Julia Thomas argues that by extending the New Keynesian model to reintroduce money's liquidity role, we can resolve some of the remaining divorce between economic theory and the patterns observed in the workings of actual economies.

Each meeting of the FOMC is met with widespread interest by everyone from financial market participants on Wall Street, to real estate agents, to the cashier at your local grocery store. People perceive changes in the FOMC's target for the federal funds rate — the interest rate at which banks borrow and lend to each other, usually overnight, through the federal funds market — as relevant and important in their everyday lives. Business people view changes in this interest rate as an important determinant influencing everything from car and home sales to consumer spending over the Christmas holiday season. Whenever business conditions are widely perceived to be weak, most people welcome cuts in the federal funds rate.

Despite these observations, however, the means through which changes in an interest rate affect business activity is, in fact, far from obvious. Over the past few decades, economists have devoted ever-growing effort to developing formal economic models to help us understand precisely how changes in interest rates brought about by monetary policy actions affect the production and provision of goods and services throughout the economy. While there are several different types of models describing how monetary policy actions drive short-run changes in total employment and GDP, a growing consensus has emerged. Most often, when an economic model is used as an additional tool with which to analyze the consequences of alternative monetary policy actions, it is drawn from a class of models known as New Keynesian (or sticky price) models.

New Keynesian models have broad appeal because they provide a relatively simple explanation for how changes in the stock of money can affect business activity and because they are, in some respects, quite consistent with what economists know about how actual changes in the money stock affect the economy. Unfortunately, though, versions of these models capable of generating realistic effects of changes in the money stock for production and employment are, at their most basic level, inconsistent with what we know

Julia Thomas is an associate professor of economics at Ohio State University. When she wrote this article she was an economic advisor and economist in the Philadelphia Fed's Research Department. This article is available free of charge at www.philadelphiafed.org/research-and-data/publications/.

*The views expressed here are those of the author and do not necessarily represent the views of the Federal Reserve Bank of Philadelphia or the Federal Reserve System.

about how interest rates move with policy-induced changes in the stock of money.

This article argues that, by extending the New Keynesian model to reintroduce an abandoned *liquidity* role of money found in earlier models, we can resolve some of the remaining divorce between our economic theory and the patterns we observe in the workings of actual economies.[1] What is this role of money? It is the idea, from classical economics, that money serves a special purpose in allowing transactions to take place between buyers and sellers, since it is the only financial asset universally accepted as a means of payment. Other assets, such as stocks and bonds, are typically not accepted as a means of payment and cannot be directly used to buy goods and services. Thus, in contrast to money, these nonmonetary assets are relatively illiquid.

When we introduce the classic liquidity role of money into the New Keynesian model, and we acknowledge the fact that it is costly to convert nonmonetary assets into monetary ones (and vice versa), we arrive at a richer model that is consistent with our knowledge of how interest rates are affected by changes in the stock of money. At the same time, the mechanics of the New Keynesian model become more complicated with this improvement, because the level of an individual's monetary assets takes on an independent role in his or her spending decisions. Exploring the effects of changes in monetary policy in this richer environment, we find that the overall magnitude of these effects and the rate at which they spread throughout the economy can depend importantly on how much money is typically held and how rapidly it changes hands, on average. In short, our extended theoretical model offers new insights about how the effects of monetary policy are transmitted throughout the economy.

Macroeconomists generally associate an easing of monetary policy with a cut in interest rates.

WHAT HAPPENS FOLLOWING A CHANGE IN MONETARY POLICY?

For many economists, at the most basic level, the changes in the economy associated with a change in monetary policy may be traced to changes in the rate at which the supply of money grows over time, rather than to movements in the interest rate. Indeed, the means through which central banks actually move their key interest rates is through open market operations, wherein government bonds — a nonmonetary asset — are exchanged for money. For example, the monetary authority can reduce the overall level of money in the economy by undertaking an open market sale of government bonds for money.[2] In the process of such a contractionary open market operation, the overall supply of bonds for sale is increased, which puts downward pressure on the price at which each bond is sold. This, in turn, increases the difference between a bond's payoff at maturity (its par value) relative to its purchase price today, ultimately raising the rate of return on bonds — that is, the interest rate.

Macroeconomists generally associate an easing of monetary policy with a cut in interest rates. As Nobel Laureate Milton Friedman put it in his 1968 presidential address to the American Economic Association, "The initial impact of increasing the quantity of money at a faster rate than it has been increasing is to make interest rates lower for a time than they would otherwise have been." Indeed, there is such consensus about the inverse relationship between short-term interest rates and the growth rate of the aggregate money supply that the relationship has been given a name: the *liquidity effect*.[3]

There is even greater consensus that changes in nominal variables, such as the interest rate, have notable consequences for the paths of real

[1] The expanded model we pursue throughout this discussion is drawn from my article with Robert King, which builds upon my work with Aubhik Khan.

[2] See the article by Frederic Mishkin or Dean Croushore's book for a more thorough discussion of the implementation of open market operations.

[3] Most evidence of the liquidity effect is indirect, in that the relationship is inferred by examining economic data through the lens of complex empirical models beyond the scope of this article. However, Seth Carpenter and Selva Demiralp directly establish the existence of the liquidity effect at a daily frequency by studying the forecast errors made at the New York Fed's Trading Desk in conducting open market operations on behalf of the Federal Reserve System. Using these errors to identify exogenous changes in the supply of reserves to the banking system, the authors establish a negative and statistically significant correlation between unanticipated changes in high-powered money and the federal funds rate. Elsewhere, John Cochrane provides direct evidence that the liquidity effect exists for broader measures of money and interest rates. He examines changes in the growth of M1 (total currency and checkable deposits) and in the nominal yields on U.S. Treasuries between October 1979 and November 1982 (a historical episode throughout which the Federal Reserve expressly targeted the quantity of money held by commercial banks). Cochrane uncovers statistically significant negative effects of M1 growth on both three-month Treasury bill rates and 20-year Treasury bond rates and thereby establishes that increases in the rate of money growth are associated with declines in nominal interest rates lasting up to one year.

variables like GDP and employment. Such real effects arising from a change in monetary policy are termed nonneutralities. Perhaps the most celebrated example of nonneutrality is the observation that reductions in inflation caused by contractionary monetary policy are associated with temporary increases in unemployment, a relationship termed the Phillips curve tradeoff.[4]

NEW KEYNESIAN MODELS

It is not easy to reproduce the patterns in the movements of money, interest rates, employment, and output observed in actual economies within our economic models; however, doing so is an important step toward understanding why these patterns arise and how they may be influenced by monetary policy. To generate nonneutralities in our models, we must first find a way to overcome their tendency to exhibit a related, and quite opposite, phenomenon known as the neutrality of money. This term applies whenever changes in an economy's money stock are transmitted immediately into the overall level of prices and have no effect at all on the real quantities of goods and services produced and sold.

Neutrality of Money. To illustrate the neutrality of money, consider the following simple example of a remote island with a single good and a single currency. Let us assume that the mango is the only good valued by inhabitants of the island and that local suppliers typically harvest and sell 50 mangos each week. The single currency used to purchase these goods is the seashell; that is, islanders buy and sell mangos using only seashells. There are 100 seashells on the island this week, as in many previous weeks, and all mangos are sold (and all seashells are exchanged for mangos) precisely once each week. Under these circumstances, the price of a mango will be two seashells.

Next, let us suppose that a nearby hurricane causes 100 additional seashells to wash up on the island's beaches next week, unexpectedly doubling the supply of currency (or money). This would seem to imply twice as many island dollars next week chasing after the same weekly harvest of 50 mangos. So what will happen to the price of a mango? One possibility is that it will immediately rise to four seashells, thereby doubling the island price level, with no change in the number of mangos harvested and sold. If this happens, the rise in the money supply will have simply led to a proportionate rise in the price level, with no consequence at all for the island's real activity – its employment and GDP – and we have a textbook case of the neutrality of money.

Nonneutrality of Money. In contrast to the scenario suggested above, most economists are convinced that actual economies exhibit short-term departures from the neutrality of money. Like the many individuals actively watching for the outcome of each meeting of the FOMC, economists generally accept that changes in the supply of money induce temporary movements in output, employment, and the real return to holding assets measured in units of consumption — the *real* (or inflation-adjusted) interest rate, to which we will return later in this article.[5] Let us reconsider our island economy of seashells and mangos in light of this consensus view.

It is not easy to reproduce the patterns in the movements of money, interest rates, employment, and output observed in actual economies within our economic models.

If the price of each mango does not immediately double in response to the unexpected doubling of seashells on the island, the quantity of mangos supplied must rise to prevent unfilled demand for mangos and undesired idle seashells. But how might this happen? New Keynesian models have a simple answer to the question. They assume that the firms supplying goods and services — in our example, the islanders gathering mangos — cannot

[4] This relationship is named after Alban William Phillips, who documented an inverse relationship between changes in unemployment and nominal wages in the United Kingdom across roughly 100 years of data. However, some argue that acknowledgment should instead go to Irving Fisher, who had suggested a similar relationship roughly 20 years earlier. The relationship was theoretically formalized to consider its policy implications by Paul Samuelson and Robert Solow.

[5] Economists use rich empirical methods to study the joint movement of interest rates, prices, and output. Their findings suggest that a persistent increase in the nominal interest rate is initially accompanied by a small decline in the growth rate of output, with little or no change in the growth rate of the price level. Over the course of several subsequent quarters, it is followed by declines in both output growth and inflation. At some point thereafter, the changes in the quantities of goods and services produced in response to the change in the interest rate eventually vanish. For further discussion, see the articles by Lawrence Christiano, Martin Eichenbaum, and Charles Evans; Harald Uhlig; and Robert King and Mark Watson.

always change their prices at will. Rather, some must honor prices that they set in the past.[6]

For simplicity, suppose that one-third of the mango sellers on our island are able to change their prices in any given week, with a single crayon used to reset prices on cardboard advertisements alternating between each of the three groups of sellers on the island each week. In this case, when the new seashells arrive, the average price of a mango will not immediately jump to four seashells. Instead, the island price level will rise only part way in the first week, since only one-third of all sellers can respond to the increase in the supply of seashells with an increase in their prices.

Assuming that all sellers are forced to supply the quantity of mangos that is demanded of them at their posted prices (or that they face sufficiently harsh penalties for not doing so that they choose to comply), the staggered price adjustment described above is all that is needed to break the neutrality of money in our island economy. With the average price in the economy not initially doubling, and assuming that all consumers on the island spend their extra seashells, the total demand for mangos will rise above the usual weekly supply of 50, and more mangos will have to be harvested. As a result, mango suppliers (and their employees) will work more relative to the normal level of labor effort on the island, and more fruit will be sold. Put another way, given temporary price stickiness among a fraction of sellers, the unexpected increase in the amount of currency on the island will have real effects, raising employment and/or the average hours worked per employee, as well as total production (real GDP).

The real effects of the rise in the island's money supply are not permanent, however. Instead, the initial week's high level of real activity will begin to subside as the economy's price level continues responding to the doubled supply of seashells. In the following week, as an additional one-third of sellers are able to raise their prices, the average price of a mango will rise further. Thus, while total demand will remain higher than usual, it will be less so than initially, and total mango production and sales will move nearer to their customary level. Eventually, as all sellers have had the opportunity to respond to the new economic conditions, the island price level will reach precisely double its original level, and the quantity of mangos harvested each week will return to the same 50 as existed before the hurricane.

The example above illustrates how unexpected increases in the money supply can temporarily stimulate economic activity. However, its mechanics are very different from the way we usually think of a change in monetary policy. Note, in particular, that our example never even mentioned a change in the interest rate! This is where the problems begin for the basic New Keynesian model.

Interest Rate Movements. In contrast to the liquidity effect observed in actual economies, the formal relationships between money, interest rates, inflation, and output at the core of the New Keynesian model lead it to predict that the interest rate *rises* when the money supply is expanded. Why do interest rates move the wrong way in the model? To understand this, we must consider a key relationship between (nominal) interest rates and *real* interest rates: the Fisher equation, named after Irving Fisher.[7] The Fisher equation says that the interest rate — the ratio of the dollar payoff on an asset relative to its dollar purchase price — is approximately equal to the sum of the real interest rate and the expected rate of inflation. To see why it is natural that this equation should hold, at least approximately, we begin with a broad definition of the real interest rate. The real interest rate is the ratio of an asset's payoff in units of future consumption of goods and services relative to the consumption

In contrast to the liquidity effect observed in actual economies, the formal relationships between money, interest rates, inflation, and output at the core of the New Keynesian model lead it to predict that the interest rate *rises* when the money supply is expanded.

[6] See the articles by William Kerr and Robert King; Bennett McCallum and Edward Nelson; and Michael Woodford for analytically tractable examples of the basic New Keynesian environment.

[7] Fisher's exposition of the relationship in *The Theory of Interest*, published in 1930, is now out of print. However, it is available online at the Library of Economics and Liberty (www.econlib.org/library/classics.html). The topic is also routinely covered in most macroeconomics texts; see, for example, Robert Barro's book.

that must be forgone today for its purchase; in other words, it is the return on savings measured not in money but in goods and services.

Returning to the island analogy above, let us suppose that our islanders are able to borrow and save. In particular, if an inhabitant saves 10 seashells this week, an island banker will lend them to some other islander and return to the original lender 11 seashells next week. Thus, the weekly nominal interest rate is 10 percent. Suppose also that the price of a mango will rise over the course of the week from one seashell to 1.01 seashells; in other words, the weekly rate of inflation is 1 percent. Under these circumstances, a mango forgone this week implies one seashell of savings deposited with the island banker that will return 1.1 seashells next week (each worth 1/1.01 mangos), allowing the lender to buy 1.089 additional mangos at that time. Notice that the real interest rate, measured in units of island goods, is then approximately 9 percent. In this way, we have arrived at the key relationship defined by Irving Fisher; the interest rate on our island is roughly equal to the sum of the real interest rate and the inflation rate.

Given the discussion above, it is straightforward to summarize why the basic New Keynesian model fares poorly with regard to the liquidity effect. In the basic model economy, an increase in the money supply implies very little change in the real interest rate. However, at the same time, it leads to comparatively substantial increases in future inflation rates. Referring back to the Fisher equation, it is then natural that the model should predict that the interest rate initially rises when the supply of money in the economy is expanded. This is a somewhat disconcerting feature of our standard model, given the broad consensus regarding the liquidity effect — the inverse relationship between changes in interest rates and changes in the money supply observed in actual economies.

EXTENDING THE MODEL: TRANSACTIONS ROLE FOR MONEY AND INFREQUENT PORTFOLIO ADJUSTMENTS

Basic New Keynesian models fail to reproduce the liquidity effect essentially because they place no emphasis on the nature of the open market operations that implement monetary policy.[8] We can correct this problem if we extend our theoretical model to reflect the fact that individuals hold both liquid assets, broadly interpretable as money, as well as illiquid assets, such as stocks and government bonds, and we also take account of the fact that individuals infrequently adjust their portfolios between these two types of assets.

In this extension of the model, we allow money to serve a particular purpose not reflected in the basic New Keynesian environment. Here, we acknowledge the fact that individuals hold low-yield liquid assets, or money, because they must draw on them for transactions. Quite simply, goods and services can be purchased only with money (which we might think of as currency, checkable deposits, and time and savings deposits). At the same time, individuals also choose to hold higher-yield nonmonetary assets, such as government bonds, as a means of saving. While these assets cannot be used directly for transactions, they pay significantly higher rates of return than money.

Various events — some expected, some unexpected — occasionally lead people to adjust their asset portfolios, moving wealth out of bonds (illiquid assets) into money (liquid assets), or vice versa. When an individual puts a down payment on a mortgage, she may do so by converting CDs or other high-yield assets into money that is deposited into her bank account and then write a check from that account to make the down payment. However, for the average person, such events are relatively infrequent. Thus, in any given month, most individuals are not actively adjusting their asset portfolios — or what we will loosely term "active in the bond market."

When there is a change in the quantity of bonds in the economy, it affects those people who are active in the bond market at that time, whether directly or through their brokers. It is with these individuals

When there is a change in the quantity of bonds in the economy, it affects those people who are active in the bond market at that time, whether directly or through their brokers.

[8] More elaborate versions of these models do succeed in generating a liquidity effect. However, Bill Dupor, Jing Han, and Yi-Chan Tsai raise an inherent tension regarding this success. They find that the additional assumptions needed to make the basic New Keynesian model consistent with the observed responses in interest rates, inflation, and output following changes in monetary policy have the unfortunate consequence of making it inconsistent with observed responses following nonmonetary disturbances.

that the monetary authority conducts an open market operation.[9] For example, the monetary authority might repurchase bonds from them and pay for the bonds by making deposits into their bank accounts. When these individuals are induced to sell bonds and receive the associated payments of money into their bank accounts, the overall supply of money in the economy is increased. However, the full rise in the stock of money does not find its way into economic activity right away. Instead, much of it remains in the recipients' bank accounts for some time.

It is precisely the fact that most people are active in the bond market only occasionally in our extended model that implies that a change in the overall money supply is not immediately transmitted throughout the economy. Most of the individuals involved in the expansionary open market operation from above do not expect to sell more bonds in the near future, so they save much of the current increase in their bank accounts to finance their expenditures over future months and boost their spending only gradually. Thus, the injection of new money into the economy does not lead to an immediate equivalent increase in aggregate spending but instead induces a more protracted rise in spending as more and more of the additional money is drawn from the recipients' accounts.

The slow increase in overall nominal spending in our extended model reduces the upward pressure on inflation relative to that in the basic New Keynesian model. How might this alter the model's performance with regard to the liquidity effect? Recalling the Fisher relationship from above, we know that the more gradual rise in inflation increases the likelihood that the interest rate will fall in response to a money injection. All that is required for this to happen is that the real interest rate exhibit a fall of sufficient magnitude to outweigh the initial rise in inflation.

This brings us to the fall in the real interest rate. For the increase in the money supply to find its way into general economic activity, individuals participating in the open market operation must be induced to increase their spending and thus their real consumption of goods and services. This can only happen, however, if the opportunity cost of an increase in their current consumption (the forgone return of a greater increase in consumption next month) is not intolerably high. To ensure that this is the case, the real interest rate must fall relative to its average level, which is precisely what happens in our extended version of the New Keynesian model. On balance, our extended model delivers a fall in the real interest rate while simultaneously reducing the upward pressure on inflation, and thus it has the ability to reproduce the liquidity effects we see following expansionary open market operations in actual economies.

MONEY VELOCITY IN THE EXTENDED MODEL

To reconcile reductions in short-term nominal interest rates with expansionary monetary policy that stimulates output and employment over the short run, we have extended the New Keynesian model to introduce an explicit transactions role for money, alongside infrequent trading of bonds by the typical individual. However, the repercussions of this extension go beyond merely resolving the problem of the absent liquidity effect. In fact, the new elements we have introduced into the model can have large and important implications for the way in which monetary policy affects the economy, because they, in turn, create a prominent role for movements in the velocity of money.

Velocity Defined. The velocity of money is another classic feature of models of the monetary economy that has been largely ignored in New Keynesian models. It is a very basic concept reflecting the average number

To reconcile reductions in short-term nominal interest rates with expansionary monetary policy that stimulates output and employment over the short run, we have extended the New Keynesian model to introduce an explicit transactions role for money, alongside infrequent trading of bonds by the typical individual.

[9] For expositional convenience, we proceed through the remainder of this discussion as though the monetary authority directly interacts with individuals when conducting open market operations. In reality, of course, interactions between the Federal Reserve System and individuals are not direct, since the Desk actually conducts open market operations through the primary dealers.

of times a unit of money is used within a specific time period, and it lies at the heart of traditional monetary theory. To compute velocity, we need only take the ratio of total nominal spending on goods and services relative to the overall stock of money in the economy. This observation comes straight from the velocity equation MV = PY, wherein M represents the aggregate money stock, V is velocity, P is the aggregate price level, and Y is real aggregate output. Notice that by simply rearranging the velocity equation, we have V = PY/M.

Let us consider our island economy once again. There, within a typical week, all seashells changed hands exactly one time, with a total of 100 available seashells being used to buy 100 seashells' worth of mangos. Thus, the weekly velocity of money was one. Now, let us suppose that, when the extra 100 seashells wash onto the island in the week of the hurricane, only one person is out on the beach to receive the unexpected "money injection," so that he is the only inhabitant to receive any additional money or even know of it. If we further suppose that this islander spends only 50 of the extra seashells this week and tucks the remainder away for future use (holding them idle in his hut for quick and costless access), total nominal spending on the island will rise to only 150 seashells out of a total seashell supply of 200. Thus, the average number of times any one seashell changes hands in the week will be 150/200, implying a money velocity of 0.75.

In our example above, when only one islander was on the beach to receive the unexpected injection of seashells, and he chose to hold half of the injection idle rather than immediately spending it or investing it in island bonds, we saw that the velocity of money dropped from its average weekly level of 1 to 0.75. This is analogous to what happens in our extended version of the New Keynesian model following an expansionary open market operation. Because only a fraction of all individuals actually take part in the open market operation, and those individuals that do participate elect to save much of the increased money stock in their bank accounts to finance near-term expenditures, there too velocity falls with an increase in the money supply.

> When an open market operation increases the bank balances of individuals who are trading bonds, their spending rises, but it rises by less than the increase in their bank accounts.

How Changes in Velocity Influence the Transmission of Monetary Policy. Changes in velocity over time can have important consequences for the rate at which nominal phenomena, such as unexpected movements in the supply of money, transmit themselves into real effects. In the basic New Keynesian model, where money has no distinct role in facilitating transactions, movements in velocity do not feed back into the operation of the real economy. It is true that money helps to determine the interest rate through the interaction of money demand and the aggregate money supply. However, once the interest rate is determined, the aggregate quantity of money and the velocity of money have no further role.[10] Put another way, changes in interest rates always affect output, employment, and inflation in the same way, irrespective of the money supply and the resulting number of times each currency unit is used.

In our expanded model, by contrast, individuals' bank balances help determine their spending over and above their total income or wealth. An individual with a total wealth of $1000, but with only $100 currently available as money in her bank account, will spend less on nondurable goods this week than will another individual who has the same $1000 but who holds it entirely in her bank account. Because money is necessary for transactions in our expanded model economy, the role of the aggregate money stock and its velocity does not end with the interest rate. Rather, the quantity of money that individuals hold and the rates at which they spend it have a direct influence on the aggregate demand for goods and services even after the interest rate has been determined. Thus, we cannot anticipate the changes in production, employment, and inflation that will follow a given change in monetary policy by simply knowing the implied path of interest

[10] This is essentially because the system of equations governing the model has only a single equation involving the demand for real balances, and that equation is effectively quarantined from the rest of the economy in that it links real balances only to the nominal interest rate and the money growth rate. Apart from the money demand equation, there is a core block of equations that contain no monetary variables at all but that together determine output, inflation, and the real interest rate as a function of the interest rate. In the most basic formulation of the model, this block of equations is simply (1) an Euler equation describing households' optimal savings behavior, (2) the Fisher relation discussed above, and (3) a Phillips curve relating current inflation and the aggregate supply of goods and services to expected inflation.

rates; instead, we must also know how individuals' money holdings and their money spending rates (velocities) will respond to the change in policy.

When an open market operation increases the bank balances of individuals who are trading bonds, their spending rises, but it rises by less than the increase in their bank accounts. Thus, we see a rise in the fraction of the money supply sitting idle awaiting future use, money changes hands less frequently than before, and velocity falls. Unlike the basic New Keynesian model, where changes in velocity have no independent influence on the economy, the decline in velocity in our expanded model has an important role in shaping the economy's response to the expansion of the money supply. When velocity falls, there are fewer dollars in circulation for undertaking transactions than there would be otherwise. This places a restraint on the economy's overall demand for goods and services and thus dampens the initial rise in production and employment. Moreover, recalling our money velocity equation from above, we know that the fall in velocity (V) means that aggregate nominal spending (PY) initially rises by less than the rise in the money supply (M).

Thus, the fall in velocity helps to restrain the rise in the aggregate price level, and the inflation rate rises by less than it would were velocity unchanged or irrelevant (as in the basic model).

Over time, as the individuals who participated in the open market operation begin to spend more and more of the extra money they are holding, aggregate velocity begins to rise back toward its normal level. Over the early part of this transition, as more and more money balances enter circulation, aggregate demand continues to rise, thereby propping up the responses in employment and output. At the same time, the rises in aggregate nominal spending must also serve to prop up the inflation rate. Thus, we see that, while the initial decline in velocity dampens the initial changes in both real quantities and inflation, these subsequent upward movements in velocity serve to protract those changes. For this reason, our economy's responses to an open market operation cannot be completed until velocity has recovered to its normal level, when the full increase in the money supply has found its way into circulation and individuals have resumed their usual spending rates.

As indicated above, the time it takes for an open market operation to flow throughout our model economy will depend on how long it takes for velocity to return to its ordinary level. In a setting where velocity is initially very high, money changes hands very frequently. There, despite some resulting decline in velocity as described above, the effects of a change in monetary policy that are unique to our expanded model are likely to vanish rapidly. This is because new money held by individuals participating in an open market operation will not be left idle for long but will instead rapidly enter circulation. After that has happened, the aggregate responses in our expanded model economy will closely resemble those of the basic New Keynesian model. By contrast, a setting with low initial velocity is one where people spend their money slowly, and the fall in aggregate velocity following an expansionary open market operation only reinforces this fact. In that setting, it will take far longer for the full effects of the same increase in the aggregate money supply to be transmitted through the economy, since it will take far more time for the new balances to fully enter into circulation.

The movements in velocity arising in our expanded New Keynesian model are, in truth, an attempt to formalize Milton Friedman's views on the transmission of monetary policy. In his words, "The initial effect of a change in monetary growth is an offsetting movement in velocity, followed by changes in the growth of spending initially manifested in output and employment, and only later in inflation."[11] If the nominal interest rate is cut when velocity is low, we will observe a slow and gradual response in output, employment, and prices in our model economy. However, the transmission of an expansionary change in policy will look quite different if velocity is high. In that case, the increase in money supply corresponding to a nominal interest rate cut will quickly find its way into circulation, yielding a more abrupt rise in production and employment and more quickly bringing about the full

When velocity falls, there are fewer dollars in circulation for undertaking transactions than there would be otherwise.

[11] This passage is drawn from Friedman's testimony to the House of Commons Select Committee in 1979; for the full text, see the 1980 reference to Friedman.

implied rise in inflation.

By extending the New Keynesian model to correct its prediction regarding the liquidity effect, we have arrived at a richer setting where movements in the velocity of money over time themselves feed back through the economy to influence how much and for how long changes in monetary policy affect real activity. As a result, our expanded theory suggests that central bankers must be attentive to more than just the change in the nominal interest rate and a simple Phillips curve relationship in considering the effects of a change in policy. They must also take into account the ways in which velocity will affect the transmission of monetary policy. Since velocity is, in part, determined by individuals' bank account balances, these balances become relevant as we anticipate the consequences of a policy change. Moreover, our theory suggests that we need to know something about individuals' willingness to alter their money spending rates over time, since this too will influence how velocity responds to a change in the growth rate of the money supply.

CONCLUSION

Economists use New Keynesian models to study how short-term nonneutralities allow monetary policy to affect real economic activity. The basic New Keynesian model explains how changes in money supply can yield temporary changes in output and employment. However, it does not explain why nominal interest rates fall when the central bank increases the money supply through an open market operation. We have discussed an extension of the model that corrects this problem by introducing an explicit transactions role for money and taking into account the fact that individuals adjust their portfolios of bonds and money infrequently.

This more complex model reconciling the New Keynesian theory with a liquidity effect exhibits important changes in the velocity of money over time. These changes vary from one economy to another depending on how much money individuals need to hold against their coming spending and depending on how willing they are to alter their money savings patterns in response to changes in aggregate conditions. Our theory predicts that the effects of changes in monetary policy will depend on both the average velocity in an economy as well as its movements over time. Thus, to anticipate the effects of a particular change in policy, we need to be able to predict how velocity will evolve in response to the change in the nominal interest rate.

On balance, when we extend our standard model to achieve greater realism with regard to interest rate movements, we find that monetary policy becomes a more complicated exercise than we may have thought and that it cannot be well understood without explicit attention to the determinants underlying the overall demand for money balances throughout the economy.

REFERENCES

Alvarez, Fernando, Andrew Atkeson, and Patrick Kehoe. "Money, Interest Rates and Exchange Rates with Endogenously Segmented Markets," *Journal of Political Economy*, 110:1 (2002), pp. 73-112.

Barro, Robert J. *Macroeconomics, 5th Edition*. Cambridge, MA: The MIT Press, 1997.

Carpenter, Seth, and Selva Demiralp. "The Liquidity Effect in the Federal Funds Market: Evidence from Daily Open Market Operations," *Journal of Money, Credit, and Banking,* 38:4 (2006), pp. 901-20.

Christiano, Lawrence J., Martin Eichenbaum, and Charles L. Evans. "Monetary Policy Shocks: What Have We Learned and to What End?" In Michael Woodford and John B. Taylor, eds., *Handbook of Macroeconomics IA*. Amsterdam: Elsevier Science, 1999.

Cochrane, John H. "The Return of the Liquidity Effect: A Study of the Short-Run Relation between Money Growth and Interest Rates," *Journal of Business and Economic Statistics,* 7:1 (1989), pp. 75-83.

Croushore, Dean. *Money and Banking: A Policy-Oriented Approach*. Boston: Houghton Mifflin Company, 2006.

Dupor, Bill, Jing Han, and Yi-Chan Tsai. "What Do Technology Shocks Tell Us About the New Keynesian Paradigm?," Ohio State University Working Paper (2007).

Friedman, Milton. "The Role of Monetary Policy," *American Economic Review,* 58:1 (1968), pp. 1-17.

Friedman, Milton. "Memorandum to U.K. Treasury and Civil Service Committee Regarding 'Enquiry into Monetary Policy'," House of Commons, United Kingdom (July 1980).

Kerr, William, and Robert G. King. "Limits on Interest Rules in the IS Model," Federal Reserve Bank of Richmond *Quarterly Review*, 82:2 (1996), pp. 47-75.

Khan, Aubhik, and Julia K. Thomas. "Inflation and Interest Rates with Endogenous Market Segmentation," Federal Reserve Bank of Philadelphia Working Paper 07-1 (2007).

King, Robert G., and Julia K. Thomas. "Breaking the New Keynesian Dichotomy: Asset Market Segmentation and the Monetary Transmission Mechanism," Working Paper (2007).

King, Robert G., and Mark W. Watson. "Money, Prices, Interest Rates and the Business Cycle," *Review of Economics and Statistics*, 78:1 (1996), pp. 35-53.

McCallum, Bennett T., and Edward M. Nelson. "An Optimizing IS-LM Specification for Monetary Policy and Business Cycle Analysis," *Journal of Money, Credit, and Banking,* 31:3 (1999), pp. 296-316.

Mishkin, Frederic S. *The Economics of Money, Banking, and Financial Markets, 8th edition*. Addison Wesley, 2006.

Phillips, A. W. (1958). "The Relationship between Unemployment and the Rate of Change of Money Wages in the United Kingdom 1861-1957," *Economica,* 25 (1958), pp. 283-99.

Samuelson, Paul A., and Robert M. Solow. "Analytical Aspects of Anti-Inflation Policy," *American Economic Review Papers and Proceedings,* 50:2 (1960), pp. 177-94.

Uhlig, Harald. "What Are the Effects of Monetary Policy on Output? Results from an Agnostic Identification Procedure," *Journal of Monetary Economics,* 52:2, pp. 381-419 (March 2005).

Woodford, Michael. "How Important Is Money in the Conduct of Monetary Policy?," NBER Working Paper 13335 (2007).

RESEARCH RAP

Abstracts of research papers produced by the economists at the Philadelphia Fed

You can find more Research Rap abstracts on our website at: www.philadelphiafed.org/research-and-data/publications/research-rap/. Or view our working papers at: www.philadelphiafed.org/research-and-data/publications/.

FIRM DYNAMICS, PRIVATE INFORMATION, AND THE GENERATION OF NEW TECHNOLOGY

The authors present a theory of spinoffs in which the key ingredient is the originator's private information concerning the quality of his new idea. Because quality is privately observed, by the standard adverse-selection logic, the market can at best offer a price that reflects the average quality of ideas sold. This gives the holders of above-average-quality ideas the incentive to spin off. The authors show that only workers with very good ideas decide to spin off, while workers with mediocre ideas sell them. Entrepreneurs of existing firms pay a price for the ideas sold in the market that implies zero expected profits for them. Hence, firms' project selection is independent of firm size, which, under some additional assumptions, leads to scale-independent growth. The entry and growth process of firms leads to invariant firm-size distributions that resemble the ones for the U.S. economy and most of its individual industries.

Working Paper 08-26, "Spinoffs and the Market for Ideas," Satyajit Chatterjee, Federal Reserve Bank of Philadelphia, and Esteban Rossi-Hansberg, Princeton University, and Visiting Scholar, Federal Reserve Bank of Philadelphia

TESTING FOR DATA RATIONALITY

Rationality of early release data is typically tested using linear regressions. Thus, failure to reject the null does not rule out the possibility of nonlinear dependence. This paper proposes two tests that instead have power against generic nonlinear alternatives. A Monte Carlo study shows that the suggested tests have good finite sample properties. Additionally, the authors carry out an empirical illustration using a real-time data set for money, output, and prices. Overall, they find strong evidence against data rationality. Interestingly, for money stock, the null is not rejected by linear tests but is rejected by the authors' tests.

Working Paper 08-27, "Information in the Revision Process of Real-Time Data Sets," Valentina Corradi, University of Warwick; Andres Fernandez, Rutgers University and Universidad de Los Andes; and Norman Swanson, Rutgers University, and Visiting Scholar, Federal Reserve Bank of Philadelphia

NONRESPONSE BIAS IN CPI MEASURES FOR RENTS

Until the end of 1977, the U.S. consumer price index for rents tended to omit rent increases when units had a change of tenants or were vacant, biasing inflation estimates downward. Beginning in 1978, the Bureau of Labor Statistics (BLS) implemented a series of methodological changes that reduced this nonresponse bias, but substantial bias remained until 1985. The authors set up a model of nonresponse bias, parameterize it, and test it

using a BLS micro-data set for rents. From 1940 to 1985, the official BLS CPI-W price index for tenant rents rose 3.6 percent annually; the authors argue that it should have risen 5.0 percent annually. Rents in 1940 should be only half as much as their official relative price; this has important consequences for historical measures of rent-house-price ratios and for the growth of real consumption. (Revision forthcoming in *Review of Economics and Statistics*.)

Working Paper 08-28, "Rents Have Been Rising, Not Falling, in the Postwar Period," Theodore Crone, Swarthmore College; Leonard I. Nakamura, Federal Reserve Bank of Philadelphia; and Richard Voith, Econsult Corporation

DESIGNING MONETARY POLICY FOR THE EURO AREA

In this paper, the authors aim to design a monetary policy for the euro area that is robust to the high degree of model uncertainty at the start of monetary union and allows for learning about model probabilities. To this end, they compare and ultimately combine Bayesian and worst-case analysis using four reference models estimated with pre-EMU synthetic data. The authors start by computing the cost of insurance against model uncertainty, that is, the relative performance of worst-case or minimax policy versus Bayesian policy. While maximum insurance comes at moderate costs, they highlight three shortcomings of this worst-case insurance policy: (i) prior beliefs that would rationalize it from a Bayesian perspective indicate that such insurance is strongly oriented toward the model with highest baseline losses; (ii) the minimax policy is not as tolerant of small perturbations of policy parameters as the Bayesian policy; and (iii) the minimax policy offers no avenue for incorporating posterior model probabilities derived from data available since monetary union. Thus, the authors propose preferences for robust policy design that reflect a mixture of the Bayesian and minimax approaches. They show how the incoming EMU data may then be used to update model probabilities and investigate the implications for policy.

Working Paper 08-29, "Insurance Policies for Monetary Policy in the Euro Area," Keith Kuester, Federal Reserve Bank of Philadelphia, and Volker Wieland, Goethe University of Frankfurt

CHOOSING THE OPTIMAL MONETARY POLICY INSTRUMENT

Currently there is a growing literature exploring the features of optimal monetary policy in New Keynesian models under both commitment and discretion. This literature usually solves for the optimal allocations that are consistent with a rational expectations market equilibrium, but it does not study how the policy can be implemented given the available policy instruments. Recently, however, King and Wolman (2004) have shown that a time-consistent policy cannot be implemented through the control of nominal money balances. In particular, they find that equilibria are not unique under a money stock regime. The authors of this paper find that King and Wolman's conclusion of non-uniqueness of Markov-perfect equilibria is sensitive to the instrument of choice. Surprisingly, if, instead, the monetary authority chooses the nominal interest rate, there exists a unique Markov-perfect equilibrium. The authors then investigate under what conditions a time-consistent planner can implement the optimal allocation by just announcing his policy rule in a decentralized setting.

Working Paper 08-30, "On the Implementation of Markov-Perfect Interest Rate and Money Supply Rules: Global and Local Uniqueness," Michael Dotsey, Federal Reserve Bank of Philadelphia, and Andreas Hornstein, Federal Reserve Bank of Richmond

BUSINESS CYCLE COSTS AND FLUCTUATIONS IN UNEMPLOYMENT

This paper develops a real business cycle model with labor market search and matching frictions, which endogenously links both the cyclical fluctuations and the mean level of unemployment to the aggregate business cycle risk. The key result of the paper is that business cycles are costly for all consumers, regardless of their wealth, yet that unemployment fluctuations themselves are not the source of these costs. Rather fluctuations over the cycle induce higher average unemployment rates as employment is non-linear in job-finding rates and past unemployment. The authors first show this result analytically in special cases. They then calibrate a general equilibrium model with risk-averse asset-holding and liquidity-constrained workers to U.S. data. Also

under these more general circumstances, business cycles mean higher unemployment for all workers. The ensuing costs of cycles rise further for liquidity-constrained agents when replacement rates are lower or when workers' skills depend on the length of (un)employment spells.

Working Paper 08-31, "The (Un)Importance of Unemployment Fluctuations for Welfare," Philip Jung, Mannheim University, and Keith Kuester, Federal Reserve Bank of Philadelphia

DOES RESTRICTING ACCESS TO EXPENSIVE CREDIT HARM CONSUMERS?

Many policymakers and some behavioral models hold that restricting access to expensive credit helps consumers by preventing overborrowing. The author examines some short-run effects of restricting access, using household panel survey data on payday loan users collected around the imposition of binding restrictions on payday loan terms in Oregon. The results suggest that borrowing fell in Oregon relative to Washington, with former payday loan users shifting partially into plausibly inferior substitutes. Additional evidence suggests that restricting access caused deterioration in the overall financial condition of the Oregon households. The results suggest that restricting access to expensive credit harms consumers, on average.

Working Paper 08-32, "Restricting Consumer Credit Access: Household Survey Evidence on Effects Around the Oregon Rate Cap," Jonathan Zinman, Dartmouth College, and Visiting Scholar, Federal Reserve Bank of Philadelphia

IS EFFICIENCY IMPORTANT IN UNDERSTANDING INSTITUTIONAL DEVELOPMENT?

Are efficiency considerations important for understanding differences in the development of institutions? The authors model institutional quality as the degree to which obligations associated with exchanging capital can be enforced. Establishing a positive level of enforcement requires an aggregate investment of capital that is no longer available for production. When capital endowments are more unequally distributed, the bigger dispersion in marginal products makes it optimal to invest more resources in enforcement. The optimal allocation of the institutional cost across agents is not monotonic and entails a redistribution of endowments before production begins. Investing in enforcement benefits primarily agents at the bottom of the endowment distribution and leads to a reduction in consumption and income inequality. Efficiency, redistribution, and the quality of institutions are thus intricately linked and should be studied jointly.

Working Paper 08-33, "Efficient Institutions," Thorsten Koeppl, Queen's University; Cyril Monnet, Federal Reserve Bank of Philadelphia; and Erwan Quintin, Federal Reserve Bank of Dallas

LABOR MARKETS' ROLE IN EURO AREA MONETARY POLICY

In this paper, the authors explore the role of labor markets for monetary policy in the euro area in a New Keynesian model in which labor markets are characterized by search and matching frictions. They first investigate to which extent a more flexible labor market would alter the business cycle behavior and the transmission of monetary policy. They find that while a lower degree of wage rigidity makes monetary policy more effective, i.e., a monetary policy shock transmits faster onto inflation, the importance of other labor market rigidities for the transmission of shocks is rather limited. Second, having estimated the model by Bayesian techniques, the authors analyze to which extent labor market shocks, such as disturbances in the vacancy posting process, shocks to the separation rate, and variations in bargaining power, are important determinants of business cycle fluctuations. Their results point primarily towards disturbances in the bargaining process as a significant contributor to inflation and output fluctuations. In sum, the paper supports current central bank practice which appears to put considerable effort into monitoring euro area wage dynamics and which appears to treat some of the other labor market information as less important for monetary policy.

Working Paper 09-1 "The Role of Labor Markets for Euro Area Monetary Policy," Kai Christoffel, European Central Bank, Frankfurt; Keith Kuester, Federal Reserve Bank of Philadelphia; and Tobias Linzert, European Central Bank, Frankfurt

LONG-TERM SOVEREIGN DEBT: ARGENTINA AS A TEST CASE

The authors present a novel and tractable model of long-term sovereign debt. They make two sets of contributions. First, on the substantive side, using Argentina as a test case they show that unlike one-period debt models, their model of long-term sovereign debt is capable of accounting for the average spread, the average default frequency, and the average debt-to-output ratio of Argentina over the 1991-2001 period without any deterioration in the model's ability to account for Argentina's cyclical facts. Using their calibrated model the authors determine what Argentina's debt, default frequency, and welfare would have been if Argentina had issued only short-term debt. Second, on the methodological side, the authors advance the theory of sovereign debt begun in Eaton and Gersovitz (1981) by establishing the existence of an equilibrium pricing function for long-term sovereign debt and by providing a fairly complete set of characterization results regarding equilibrium default and borrowing behavior. In addition, they identify and solve a computational problem associated with pricing long-term unsecured debt that stems from nonconvexities introduced by the possibility of default.

Working Paper 09-2, "Maturity, Indebtedness, and Default Risk," Satyajit Chatterjee, Federal Reserve Bank of Philadelphia, Burcu Eyigungor, Koç University

WHOLESALE FUNDS, MARKET DISCIPLINE, AND LIQUIDITY RISKS

Commercial banks increasingly use short-term wholesale funds to supplement traditional retail deposits. The existing literature mainly points to the "bright side" of wholesale funding: sophisticated financiers can monitor banks, disciplining bad ones but refinancing solvent ones. This paper models a "dark side" of wholesale funding. In an environment with a costless but imperfect signal on bank project quality (e.g., credit ratings, performance of peers), short-term wholesale financiers have lower incentives to conduct costly information acquisition and instead may withdraw based on negative but noisy public signals, triggering inefficient liquidations. The authors show that the "dark side" of wholesale funding dominates the "bright side" when bank assets are more arm's length and tradable (leading to more relevant public signals and lower liquidation costs): precisely the attributes of a banking sector with securitizations and risk transfers. The results shed light on the recent financial turmoil, explaining why some wholesale financiers did not provide market discipline ex-ante and exacerbated liquidity risks ex-post.

Working Paper 09-3, "The Dark Side of Wholesale Funding," Rocco Huang, Federal Reserve Bank of Philadelphia, and Lev Ratnovski, International Monetary Fund