Economic Review

FEDERAL RESERVE BANK OF KANSAS CITY

September/October 1988

Maintaining Central Bank Credibility

Economies of Scale and Scope At Depository Financial Institutions: A Review of the Literature

Demographic Influences on Household Growth and Housing Activity

September/October 1988, Volume 73, No. 8

The ***Economic Review*** (ISSN0161-2387) is published monthly by the Federal Reserve Bank of Kansas City, except in **July/August** and **September/October**, when it is published bi-monthly. Subscriptions and additional copies are available without charge. Send requests to the Public Affairs Department, Research Division, Federal Reserve Bank of Kansas City, 925 Grand Avenue, Kansas City, Missouri, 64198. Please include your mailing label with any address change. If any material is reproduced from this publication, please credit the source. Second class postage paid at Kansas City, Missouri. Postmaster: send address changes to ***Economic Review,*** Research Division, Federal Reserve Bank of Kansas City, 925 Grand Avenue, Kansas City, Missouri 64198.

Economic Review

FEDERAL RESERVE BANK OF KANSAS CITY

September/October 1988

Maintaining Central Bank Credibility

By Anne Sibert and Stuart E. Weiner

Central banks throughout the world have increasingly recognized the importance of maintaining public confidence in the commitment of governments to controlling inflation. This recognition has resulted in part from the worldwide inflation in the 1970s. Whether because of oil supply shocks, excessive wage demands, or unduly expansionary government policies, inflation ratcheted upward throughout the 1970s in most industrial countries. As a result of this experience, households and businesses may have become skeptical about the ability and willingness of governments to maintain a reasonably stable price level.

Recent economic research has focused on the implications of such skepticism and what can be done to keep the public's confidence in monetary policy. One major conclusion of this research is that the credibility of a central bank's commitment to price stability can be undermined by public perceptions that keeping unemployment at an unrealistically low level is an overriding goal of monetary policy. Such perceptions can lead the public to expect an inflationary monetary policy even when the central bank announces its commitment to price stability.

Moreover, the public's fear of an inflationary policy may be particularly acute if the central bank does not conduct policy according to a fixed rule but instead exercises judgment depending on a variety of economic variables. Exercising judgment in this way is often referred to as a discretionary policy procedure. A possible problem with this procedure is that what a central bank promises to do in the future may be inconsistent with what it in fact does when the time comes for it to act. For that reason, the problem arising from discretionary monetary policies is referred to as the time-inconsistency problem, which causes a related credibility problem for a central bank in convincing the public of its commitment to price stability.

Some economists have concluded from this line of reasoning that central banks should avoid discretion and be required to adopt monetary

Anne Sibert is an assistant professor of economics at the University of Kansas and a visiting scholar at the Federal Reserve Bank of Kansas City. Stuart E. Weiner is a research officer and economist at the bank.

growth rules that will relieve the public's anxiety about the prospect of inflationary monetary policy in the future. Recognizing the possible shortcomings of these monetary growth rules, other economists have proposed alternative solutions to the time-inconsistency and credibility problems faced by central banks.

This article argues that neither monetary growth rules nor other proposed solutions to the time-inconsistency problem are costless. The first section explains why discretionary conduct of monetary policy may make it difficult to establish the credibility of a central bank's commitment to price stability. The second section shows why monetary growth rules are not a costless solution to the credibility problem. And the third section demonstrates why alternative proposals are also problematical.

Time inconsistency and central bank credibility

The credibility problem arises if the public is skeptical about a central bank's intention to pursue noninflationary monetary policies. Individuals and businesses might be skeptical, for example, if they perceive that the central bank would like to lower unemployment temporarily below the rate that can be achieved in a noninflationary environment. Even if the public and the central bank understand that attempting to temporarily lower unemployment may not be successful and will result in permanently higher inflation that will damage the economy, the public may distrust a central bank's assertions that it will not succumb to this temptation to inflate. Ironically, the public's skepticism arises in large part because private citizens realize that they might well pursue such an inflationary policy if placed in the position of central bankers.[1]

The tradeoff between unemployment and inflation

Society and central banks would like to achieve both low inflation and low unemployment. Inflation is considered undesirable because it contributes to social strains by creating the perception that a market economy can lead to arbitrary and unfair redistribution of wealth. Moreover, inflation imposes real economic costs by causing an inefficient allocation of society's scarce resources.[2] Unemployment is considered undesirable because it means that an important scarce resource, labor, is being underutilized. Consequently, both unemployment and inflation reduce society's overall economic welfare.

But lowering unemployment and inflation simultaneously may not be possible in the short run. An unexpected expansionary monetary policy, for example, not only leads firms to hire more workers but also causes inflation to worsen as demand for output increases. The level of inflation is thus inversely related to the level of unemployment in the short run. This inverse relationship is embodied in the short-run Phillips curve, which shows that lowering unemployment is typically associated with higher inflation.

Over longer run periods, trying to keep unemployment low through expansionary macroeconomic policies also leads to higher inflation. Not only do prices of goods rise throughout the

[1] The seminal paper in the time-inconsistency literature is Finn Kydland and Edward Prescott, "Rules Rather Than Discretion: The Inconsistency of Optimal Plans," *Journal of Political Economy*. June 1977, pp. 473-492. The idea was popularized by Robert Barro and David Gordon in "A Positive Theory of Monetary Policy in a Natural Rate Model," *Journal of Political Economy*, August 1983, pp. 589-610. Shortcomings of this literature are examined by Torsten Persson, "Credibility of Macroeconomic Policy: An Introduction and a Broad Survey," *European Economic Review*. 1988, pp. 519-532.

[2] For further discussion of the costs of inflation see Stanley Fischer, "The Benefits of Price Stability," *Price Stability and Public Policy*, proceedings of an economic symposium sponsored by the Federal Reserve Bank of Kansas City, 1984.

economy, but wages rise as well as labor contracts are renegotiated.

The amount of unemployment that is consistent with stable inflation is called the natural level of unemployment. Some unemployment is inevitable in a market economy in which workers are free to change jobs and take time off between jobs. The natural level of unemployment is thus positive.

The natural level of **unemployment** may nonetheless be above the socially optimal level of unemployment because of labor market distortions. Income taxes and social security taxes provide an example. Such taxes reduce workers' after-tax wages. By driving a wedge between what employers pay and what employees receive, income and social security taxes keep some workers from working as much as they otherwise would. As a result of such distortions, the level of unemployment that is consistent with stable inflation in the long run may well be higher than is socially desirable in a broader sense.[3]

Central banks and society thus face the dilemma of accepting undesirably high unemployment or lowering unemployment through inflationary monetary policies. The dilemma arises because society cannot achieve both of its major macroeconomic objectives simultaneously. One or the other must be sacrificed unless labor market imperfections are eliminated.

Moreover, a central bank cannot lower unemployment by pursuing inflationary policies that are fully anticipated by firms and workers. Firms decide how many workers to hire based on the real, or inflation-adjusted, wage they must pay. Similarly, workers decide whether to take a job based on the real purchasing power of the wages they are offered. Inflation that is fully anticipated would thus not alter employment decisions. Instead, anticipated inflation merely lowers social welfare because of the associated inefficiencies without any compensating reduction in unemployment.

The only way for a central bank to lower unemployment is to pursue monetary policies that cause inflation to rise **unexpectedly**.[4] If workers and firms initially expect no inflation and enter into long-term wage contracts based on this expectation, an unexpected rise in inflation can change the real wage and thus the level of employment. To see how a central bank can lower unemployment by generating surprise inflation, consider the labor market diagram in Figure 1. The real wage is measured along the vertical axis, and the employment level is measured along the horizontal axis. The real wage is the nominal wage (W) deflated by the price level (P). The labor demand curve, D,, shows the amount of labor that firms want to hire at any given real wage. It slopes downward and to the right because firms want to hire more workers as the real wage falls, that is, as labor becomes cheaper. The labor supply curve, S,, shows the amount of labor that workers want to supply at any given real wage. It slopes upward and to the right because more individuals

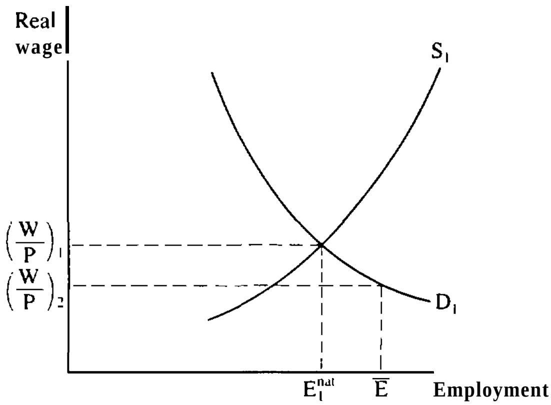
---

3 Income taxes reduce the amount of pay that workers take home but also provide revenue for government services. Given a certain tax level, workers will collectively choose to supply labor along a given labor supply curve, and overall employment will be at its natural level. Any individual worker. however, would prefer that overall employment be greater than this natural level because tax revenues, and hence government services, would be correspondingly higher. Individual workers will not increase their work effort, however, because they would view the increased government services resulting from this effort as negligible. Consequently, the higher level of "optimal" employment will not be achieved.

Other labor market features that impede the attainment of optimal employment (though not necessarily strictly within the context of the above model) include skill mismatches, location mismatches, institutional barriers, imperfect information flows, and transfer payment disincentives. For discussion, see Stuart E. Weiner, "The Natural Rate of Unemployment: Concepts and Issues," ***Economic Review***, Federal Reserve Bank of Kansas City, January 1986, pp. 11-24.

4 This description follows Kydland and Prescott, "Rules Rather Than Discretion . . . ."

FIGURE 1
The labor market



want to work as the real wage rises, that is, as workers are paid more.

Assume that workers and firms enter into a contract in which workers are paid a fixed nominal wage over a certain period of time. Workers and firms agree on this nominal wage without knowing the future price level. Once the actual price level is known, firms decide how much labor to employ based on the resulting real wage. By assumption, workers are obligated to supply this labor according to the terms of the labor agreement. Thus, the level of employment is determined by the demand for labor by firms.[5]

Suppose that the nominal wage agreed on and the associated price level yield a real wage of $(\frac{W}{P})_1$. At this real wage, the labor demand and labor supply curves intersect, so the amount of labor demanded is equal to the amount of labor supplied. Firms are on their labor demand curve, and workers are on their labor supply curve. The resulting employment level, denoted E_1^{nat}, is called the "natural level of employment." It is the employment counterpart to the natural level of unemployment discussed above.

A central bank that is allowed discretion in conducting policy can temporarily raise employment by generating surprise inflation through an unexpected increase in the money supply. Suppose that some labor market distortion, such as income taxes, causes the natural level of employment to be below the optimal level of employment (and, correspondingly, causes the natural level of unemployment to be above the optimal level of unemployment). A central bank could increase employment to the optimal level by unexpectedly pursuing an inflationary monetary policy. The unexpected rise in the price level would lower real wages because nominal wages are assumed to be fixed. In Figure 1, the real wage would decline from $(\frac{W}{P})_1$ to $(\frac{W}{P})_2$, and employment would rise from E_1^{nat} to $\bar{E}$.

The rise in employment will only be temporary, however, because workers are supplying more labor than they want to. Workers are not on their labor supply curve. If workers believe that higher prices will continue, they will compensate by negotiating higher nominal wages in the next contract negotiation. The real wage will eventually retrace its path, settling again at $(\frac{W}{P})_1$. In the end, the price level will be higher, nominal wages will be higher, and employment will be back at the natural level.

The temptation to inflate

The discussion above demonstrates that a central bank can temporarily increase employment by generating surprise inflation. This possibility causes a credibility problem for the central bank and can introduce an inflationary bias into

[5] **This assumption that employment is demand-determined is reasonable for unionized U.S. labor markets and follows that of Jo Anna Gray, "Wage Indexation: A Macroeconomic Approach,"** ***Journal of Monetary Economics,*** **April** ***1976,*** **pp. 221-235.**

monetary policy.[6] Even if a central bank announces a strict anti-inflation policy and has every intention of adhering to it, the central bank has an incentive to renege once nominal wages are fixed. This temptation to inflate once nominal wages are set is the essence of the time-inconsistency problem. Workers and firms understand the central bank's incentives and are thus skeptical about its policy announcements. Lacking credibility, the central bank is unable to increase employment and, indeed, must tolerate higher inflation to maintain existing employment.

To understand this argument, consider again the firms and workers depicted in Figure 1. Suppose the central bank announced that it would not inflate and that workers and firms negotiated a nominal wage expected to yield a real wage of $(\frac{W}{P})_1$ and a corresponding employment level of E_1^{nat}. With the nominal wage fixed, the central bank would now have an incentive to renege on its anti-inflation promise. By driving prices higher and real wages lower, the central bank is able to attain a higher level of employment, say $\bar{E}$. Workers and firms recognize this ability, however, and in fact would not agree to a nominal wage that permits it to happen. Rather, expecting inflation, workers and firms will negotiate a higher nominal wage tliat compensates for the expected inflation. The central bank will then have to inflate just to ensure that the real wage does not go ***above*** $(\frac{W}{P})_1$ and employment go ***below*** E_1^{nat}. The end result is higher inflation with no compensating reduction in unemployment.

In summary, the time-inconsistency and credibility problems arise when the public comes to doubt a central bank's commitment to price stability. The doubt arises when the public believes the central bank can and will sacrifice price stability in order to lower unemployment, even if only temporarily. A central bank can do so if it conducts monetary policy using discretion rather than rules and if real wages and thus employment are affected by unexpected inflation. A central bank may be willing to make the tradeoff if labor market distortions make it impossible to achieve the socially desirable level of employment without creating unexpected inflation. Recognizing the incentive to promise low inflation but deliver high inflation, the public could become skeptical of a central bank's commitment to price stability. Such skepticism would lead workers and firms to expect inflation in the future and seek to protect themselves by building an inflation premium into wage contracts. When this occurs, the inflation expectations become a self-fulfilling prophecy. This process may be reversed if the central bank follows a monetary policy restrictive enough to cause unemployment to remain high long enough to change the public's expectations.[7] Lack of central bank credibility can thus result in either higher inflation, higher unemployment, or both. For this reason, it is important to analyze how monetary policy can be conducted to maintain public confidence in the central bank's commitment to price stability.

Caveats

The assumption that a central bank is perceived as wanting to keep unemployment artificially low is a critical element in the argument that central banks face a serious credibility problem. If,

6 A government might also be templed to inflate for reasons other than generating employment gains. See, for example, Kenneth Rogoff and Anne Sibert, "Elections and Macroeconomic Policy Cycles," *Review of Economic Studies*, February 1988. pp. 1-16, and Guillermo Calvo, "On the Time Consistency of Optimal Policy in a Monetary Economy," *Econometrica*, November 1978, pp. 1411-1428.

7 This is the conclusion reached by David Backus and John Driffill, "Rational Expectations and Policy Credibility Following a Change in Regime," *Review of Economic Studies*, April 1985, pp. 211-222. If the private sector is uncertain of the preferences of the central bank, it may revise its beliefs about the credibility of the bank after observing the bank carrying out an anti-inflation policy.

instead, it is assumed that the public believes monetary policy is directed toward maintaining price stability and keeping the unemployment rate near the natural level, the conclusion that central banks face a serious credibility problem is much weaker. Indeed, one study in the professional literature shows that the credibility problem vanishes if the public believes the central bank's objectives do not include keeping the unemployment rate below the natural level.[8]

The authors of this important study do not attempt to provide a compelling case for why a central bank would try to reduce unemployment below the natural level. Rather, they conjecture that the socially desirable unemployment level is probably below the natural level due to labor market distortions of the type described above. That the unemployment level is kept above the socially desirable level by various microeconomic distortions does not in itself prove, however, that the central bank would be perceived as trying to remedy the problem through macroeconomic policy. It might be argued, for example, that the public and the central bank would both recognize the advisability of using policies other than monetary expansion to compensate for distortions in labor markets.

The extent to which central banks do try to keep unemployment below the natural level may well vary. The institutional arrangements of and legislative mandates for central banks vary widely. Such variation may lead some central banks to have different priorities than others. Moreover, the central bank of any particular country may emphasize certain goals more in some circumstances than in others, leading the public's skepticism about the central bank's commitment to control inflation to vary accordingly.

For these and other reasons, the conclusions of the time-inconsistency literature have by no means been universally accepted. No consensus has emerged on the practical importance of the time-inconsistency problem in explaining inflation in industrial countries. In addition, some ambiguities remain in the fundamental analysis, and several theoretical issues are unresolved.[9] Nevertheless, this literature does provide insight into the potential inflationary bias of a society and its central bank. And it provides insight into the importance of credibility.

Monetary growth rules

The principal conclusion of the central bank credibility literature is that central banks will tend

[8] See Barro and Gordon, "A Positive Theory . . . ."

[9] One unresolved issue involves the compatibility of social welfare with individual preferences. The time-inconsistency problem assumes that social welfare depends negatively upon deviations from optimal levels of employment and inflation. Thus, social welfare must be increasing in unanticipated inflation (until the optimal level of employment is reached) and decreasing in actual inflation. However, it is not obvious that this is the case. One reason is that unanticipated inflation may be costly as well as beneficial. Suppose. for example, that suppliers see their own prices rise before they observe that the general price level has risen. Then they may incorrectly conclude that the demand for their product has increased, and they may produce more than they would if their information was perfect. Another problem is that it is not clear that economic welfare is decreasing in actual inflation. The usual arguments for why this is the case are that the tax system must be changed and individuals must hold higher money balances. However, in the above arguments, it is expected inflation rather than actual inflation that is costly. For discussion, see Robert Lucas, "Expectations and the Neutrality of Money," *Journal of Economic Theory*, April 1972, pp. 103-124, and Herschel Grossman, "A General Model of Monetary Policy, Inflation, and Reputation," mimeo, 1987.

A second unresolved issue involves labor market distonions. The root of the time-inconsistency problem is the tax-induced distortions in the labor market that keep employment below its socially optimal level. But these distortionary taxes finance public goods. Suppose that at the natural level of employment tax revenue is below the socially optimal amount. Then the government will want to increase revenues. Should it do this by increasing or decreasing employment? It may be that less employment at a higher real wage leads to increased tax revenues. Thus, the government may not wish to inflate. See Alex Cukierman and Allan Drazen, "Do Distortionary Taxes Induce Policies Biased Towards Inflation? A Macroeconomic Analysis." Tel-Aviv University, August 1986.

to adopt inflationary policies unless a way can be found to limit their discretion. A rule that limits the central bank's discretion might seem an obvious solution to the credibility problem. And, indeed, rules placing constraints on monetary growth have been proposed. Unfortunately, some of the features of such rules are themselves problematical.[10]

One proposed solution to the central bank credibility problem is for the central bank to adopt a smct constant growth rate rule. Under such a rule, the central bank would be required to keep the money supply growing at a constant rate every year. The central bank could never exercise discretion to vary this growth rate.

The principal appeal of the strict constant growth rate rule is that it does in fact solve the central bank credibility problem. Although the central bank still has an incentive to inflate when nominal wages are fixed, it can no longer act on that incentive. The central bank does not have the discretion to make policy changes. Because workers and firms know the central bank must adhere to the rule, they know that the central bank cannot generate surprise inflation. Thus, the credibility problem is solved.

The principal drawback of the strict constant growth rate rule is that it prevents a central bank from responding to various shocks that occasionally disrupt the economy. These shocks—either to the supply of goods and services (supply shocks) or to the amount of money that individuals wish to hold (money demand shocks)—lead firms to employ fewer workers, causing employment to decline below its natural level. An example of a supply shock is a drought. An example of a money demand shock is a financial crisis that increases the demand for liquid assets. (See box on page 11.) Such shocks impose a cost on an economy because they reduce employment, and adherence to a constant growth rate rule would not allow monetary policy to be eased to offset these costs. The results of a strict constant growth rate rule are summarized in row 1 of Table 1.[11]

An alternative to this strict constant growth rate rule is a more flexible rule that permits the central bank to respond to supply shocks but not to money demand shocks. Under such a rule, the central bank would be required to keep the money supply growing at a constant rate unless the economy experienced a supply shock. If a supply shock occurred, the central bank could accommodate it by increasing the rate of monetary growth. If a money demand shock occurred, in contrast, the central bank could not exercise such discretion.

Liberalizing the strict constant growth rate rule in this way does not reintroduce the credibility problem. Supply shocks such as droughts can be recognized by workers and firms. As a result, the central bank could never falsely claim that it had expanded the money supply to accommodate a supply shock when its real intention was to generate surprise inflation. It is assumed that the public can discriminate between actual and alleged supply shocks.[12] Hence, the credibility problem remains solved, and there are no costs incurred because of the inability to accommodate supply shocks. The costs of not accommodating money demand shocks remain, however. The results of this constant growth rate rule adjusted for supply shocks are summarized in row 2 of Table 1.

10 These rules are described by Matthew Canzoneri in "Monetary Policy Games and the Role of Private Information," *American Economic Review.* December 1985, pp. 1056-1070.

11 These results extend to more complicated average targeting and feedback rules. See Anne Sibert, "Notes on Time-Inconsistency," unpublished notes, 1988.

12 Supply shocks cannot always be easily identified. Productivity shocks, for example, are difficult to detect.

TABLE 1

Monetary growth rule solutions to the central bank credibility problem

Solution	Economy is shock free	Economy is subject to supply shocks	Economy is subject to both supply shocks and money demand shocks
	(1)	(2)	(3)
1. Strict constant growth rate rule	Problem solved Solution costless	Problem solved Solution costly	Problem solved Solution costly
2. Constant growth rate rule adjusted for supply shocks	Problem solved Solution costless	Problem solved Solution costless	Problem solved Solution costly
3. Constant growth rate rule adjusted for supply and money demand shocks	Problem solved Solution costless	Problem solved Solution costless	Problem remains

A third approach, of course, is to adopt a growth rate rule that permits the central bank to accommodate both supply shocks and money demand shocks. Under such a rule, the central bank would be forced to keep the money supply growing at a constant rate unless the economy experienced a supply shock or a money demand shock. The central bank would have the freedom to accommodate whatever shock occurred by altering the growth of money. The chief appeal of this rule is that it would eliminate the employment and output losses associated with not reacting to money demand shocks. The chief drawback of this rule is that the credibility problem reappears.

The credibility problem reappears because, unlike supply shocks, money demand shocks cannot typically be identified by the public.[13] An increase in the preference of individuals for more liquid assets, for example, cannot easily be inferred except from empirical estimation of money demand functions. As a result, the central bank and the public must forecast money demand. Assuming that the central bank's forecast is not publicly available, the central bank will once again have an incentive to generate surprise inflation, claiming that it expanded the money supply on the mistaken belief that money demand had increased. And awareness on the part of workers and firms of this incentive may cause them to be skeptical of the central bank's claim that its empirical estimates indicate that the money demand function has shifted.

In effect, this constant growth rate rule adjusted for supply and money demand shocks is not really a rule at all. It is rather an arrangement that permits considerable discretion to the central bank. The central bank is free to change monetary growth in response to whatever real or imagined shock. There are no effective limitations on the central bank's actions. The results of such a rule are shown in row 3 of Table 1.

13 Canzoneri makes this point in "Monetary Policy Games . . . . "

Economic Shocks

An economy is typically subjected to a variety of random shocks. Two common shocks are supply shocks and money demand shocks. How monetary policy reacts to such shocks is the focus of much attention in the credibility literature.*

A supply shock, as the name would indicate, is some development that disrupts the supply of goods and services. Examples of supply shocks include an oil embargo, a crop failure, or a decline in labor productivity. When an economy experiences a supply shock, employment typically declines. How far it declines depends in pan on how the central bank responds.

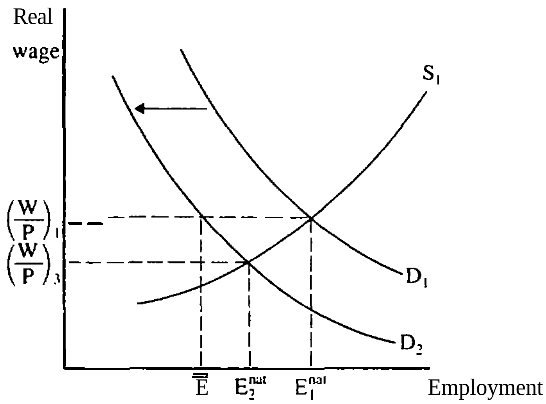
Figure 2 shows the labor market in the presence of a supply shock. Before the shock, workers and firms are assumed to be on their labor supply (S_1) and labor demand (D_1) curves, respectively, with the real wage at $(\frac{W}{P})_1$ and employment at its natural level, E_1^{nat}. Now a shock occurs—OPEC, for example, institutes an oil embargo that forces the price of oil much higher. As the price of oil rises, the demand for labor will fall because firms will want to scale back production. This decline in the demand for labor is represented in Figure 2 as a leftward shift in the labor demand curve, from D_1 to D_2.

What happens to employment? Employment falls to $\overline{\overline{E}}$, and workers are forced off their supply curve. Note that $\overline{\overline{E}}$ is below the new natural level of employment, E_2^{nat}. Employment will remain at $\overline{\overline{E}}$ until the real wage declines.

As new contracts are negotiated, the real wage will decline because workers will come to realize that nominal wage restraint is necessary if employment is to rise. However, the process could be long and hard. An alternative way to get employment at its new natural level is for the central bank to "accommodate" the supply shock. It does this by increasing the money supply, which in turn causes a rise in prices and a decline in the real wage. The right amount of accommodation will lower the real wage to $(\frac{W}{P})_3$ and raise employment to E_2^{nat}.

FIGURE 2

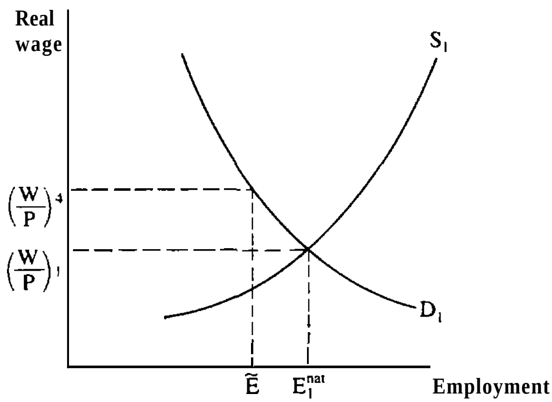
The labor market in the presence of a supply shock



A money demand shock is different. Such a shock occurs when—for reasons unrelated to changes in income or inflation—the public decides to hold more or less of its assets in the form of money.

Figure 3 shows the labor market in the presence of a money demand shock. Before the shock, workers and firms are assumed to be on their labor supply (S_1) and labor demand (D_1) curves, respectively, with the real wage at $(\frac{W}{P})_1$—and employment at its natural level, E_1^{nat}. Now a money demand shock occurs—households, for example, decide to sell some stocks and want to hold the proceeds in checking accounts. Because the supply of money has not changed, the amount of money now available to facilitate everyday transactions has declined. As a result, the price level falls and the real wage rises, to $(\frac{W}{P})_4$. At this higher real wage, firms want to hire fewer workers, so employment falls to $\tilde{E}$ and workers

FIGURE 3
The labor market in the presence of a money demand shock



are again forced off their labor supply curve. Note that $\tilde{E}$ is below the natural level of employment, E_1^{nat}. Employment will remain at $\tilde{E}$ until the real wage declines.

As in the supply shock case, when an economy experiences a money demand shock, the central bank can take steps to ensure a speedy return to the natural level of employment. Specifically, it can increase the money supply, which causes a rise in prices and a decline in the real wage. The right amount of such accommodation will lower the real wage back to $(\frac{W}{P})_1$ and raise employment back to E_1^{nat}. Thus given some discretion, the central bank can offset disruptive shocks.

***The discussion here and in the text focuses on negative shocks; that is, shocks that potentially cause output and employment losses. Positive shocks, in contrast, potentially cause output and employment gains.**

The main message of this discussion of monetary growth rules is summarized in column 3 of Table 1: In a realistic economic environment, one subject to both supply shocks and money demand shocks, either a credibility problem will remain or a solution will be costly.

Alternative solutions

Because of the problems inherent in monetary growth rules, alternative solutions to the central bank credibility problem have been proposed. These proposals do not limit the degree of central bank discretion but, rather, alter the environment in which the bank operates. Such proposals include wage indexation, a conservative central bank, and long-term relationships. A common feature of all is that they, too, are problematical.

Wage indexation

One possible solution to the credibility problem is to rely on wage indexation. Wage indexation ties nominal wages to the price level, so that nominal wages rise in line with the overall price level. Wage indexation typically takes the form of cost-of-living allowances (COLAs) in labor agreements.

Wage indexation can be either complete or partial. Complete indexation offers workers full protection against price increases: If prices rise 1 percent, nominal wages rise 1 percent, leaving the real wage unchanged. Partial indexation offers workers partial protection: If prices rise 1 percent, nominal wages rise less than 1 percent, causing some reduction in the real wage. As potential solutions to the central bank credibility problem, complete indexation and partial indexation possess different attributes.

Complete wage indexation would solve the credibility problem. A central bank would have no incentive to generate surprise inflation because expansionary monetary policy could not lower unemployment, even temporarily. Any increase in prices brought on by the central bank would be fully reflected in higher nominal wages. The

TABLE 2
Alternative solutions to the central bank credibility problem

Solution	Economy is shock free (1)	Economy is subject to supply shocks (2)	Economy is subject to both supply shocks and money demand shocks (3)
1. Complete wage indexation	Problem solved Solution costless	Problem solved Solution costly	Problem solved Solution costly
2. Partial wage indexation	Problem remains	Problem remains	Problem remains
3. Conservative central bank	Problem solved Solution costless	Problem solved Solution costly	Problem solved Solution costly
4. Long-term relationships	Problem solved Solution costless	Problem solved Solution costless	Problem solved Solution costly

real wage and the level of unemployment would remain unchanged. The central bank would thus have no incentive to pursue inflationary policies. Accordingly, the public would find a central bank's assurances of its commitment to price stability entirely credible.

Complete wage indexation would leave workers vulnerable to supply shocks, however. As noted in the previous section, a supply shock causes employment to fall below its natural level, and employment will remain below its natural level until the real wage is permitted to fall. With complete wage indexation, the real wage cannot fall. Thus, as noted in row 1 of Table 2, complete wage indexation would be costly because it would lead to higher unemployment by preventing real wages from adjusting to supply shocks.[14]

Partial wage indexation, in contrast, would allow greater flexibility of real wages but would not solve the credibility problem. Because the real wage would decline somewhat whenever prices rose, a central bank could temporarily raise employment by generating surprise inflation. Recognizing the ability of the central bank to affect real wages and thus employment, the public would be skeptical about the central bank's commitment to price stability. This point is noted in row 2 of Table 2.

Conservative central bank

A second proposed solution to the credibility problem is that of a conservative central bank. A conservative central bank can be defined as one that dislikes inflation more than society does. Such a central bank will be less inclined to generate surprise inflation in an attempt to increase employment beyond its natural level. Thus, the credibility problem will be solved. However, this solution is not costless because this same central bank might also be less willing to accommodate supply shocks by increasing the rate of monetary growth. If so, employment and output would be lost.[15]

[14] For further discussion, see Gray, "Wage Indexation . . . ."

[15] See Kenneth Rogoff, "The Optimal Degree of Commitment to an Intermediate Monetary Target," *Quarterly Journal of Economics,* November 1985, pp. 1169-1189.

This discussion is moot, however, if inflation-averse central bankers cannot be appointed. The only way to ensure that they can be appointed is to have the central bank independent of the rest of government. That is, an institutional framework needs to be established that allows the central bank to operate free of political pressure. To some extent, such a framework is in place in the United States as well as other industrialized countries. The results of this conservative central bank solution are shown in row 3 of Table 2.

Long-term relationships

A final proposed solution to the credibility problem involves long-term relationships. The central bank credibility problem might be avoided or at least reduced if the relationship between a central bank and the private sector is a lasting one. Specifically, if the actions of the central bank affect the expectations of the private sector about the future, the central bank must weigh not only the direct costs and benefits of inflation but also the impact of such inflation on inflationary expectations. If current inflation leads to a sufficient worsening of inflationary expectations, the central bank may not have an incentive to inflate in the current period.[16]

Economists have developed formal models to capture the effect of central bank actions on inflation expectations. Suppose that the private sector has the following beliefs. If the central bank has never inflated more than the socially optimal amount, excess inflation is not expected. But if the central bank ever does generate surprise inflation, excess inflation will be expected for a certain amount of time in the future. Given these beliefs, the central bank can increase employment in the short run. But the cost of doing so is higher inflation—with no accompanying employment gain—for some time in the future. If society and the central bank care enough about the future, these long-run costs may deter the central bank from generating surprise inflation.[17]

In a world free of money demand shocks, this deterrent effect of the public's inflation expectations would solve the credibility problem. Unfortunately, the world is not free of money demand shocks. Moreover, such shocks are not directly observable and, hence, must be forecasted by the central bank. As a result, when the private sector observes excess inflation, it does not know whether the central bank inadvertently inflated by incorrectly forecasting money demand or deliberately inflated in an attempt to increase employment.

But the credibility problem may still be solved, albeit at a cost. Suppose, for example, that as long as inflation remains below a certain trigger level the private sector will not expect inflation but that if inflation rises above this trigger level inflation will be expected for some time in the future. As in the previous scenario, if society and the central bank care enough about the future, the central bank may decide not to deliberately generate surprise inflation. However, because of unobserved money demand shocks, accidental inflation could arise. And if this accidental inflation exceeds the target level, the public will raise its inflation expectations. One would thus observe periods of costly excess inflation interspersed with

[16] For further discussion, see William Fellner. *Towards a Reconstruction of Macroeconomics.* American Enterprise Institute, 1976, and William Fellner, "The Credibility Effect and Rational Expectations: Implications of the Gramlich Study." *Brookings Papers on Economic Activity,* 1:1979, pp. 167-178.

[17] This model is due to Robert J. Barro and David Gordon, "Rules. Discretion, and Reputation in a Model of Monetary Policy." *Journal of Monetary Economics,* July 1983, pp. 101-121. and based on a game theory model developed by James W. Friedman, "A Noncooperative Equilibrium for Supergames," *Review of Economic Studies,* January 1971, pp. 1-12. Strictly speaking, in the Barro and Gordon model. the credibility problem is not fully solved but rather lessened.

periods of little or no inflation.[18] The characteristics of this long-term relationship are summarized in row 4 of Table 2.

This framework of a long-term relationship appears to provide some insight into developments in the United States in the 1970s and 1980s. Some economists would argue that the Federal Reserve pursued an overly expansionary monetary policy in the 1970s in an attempt to keep unemployment below its natural level. Others would argue that monetary policy was inadvertently expansionary because the Federal Reserve had imprecise information about supply shocks, money demand shocks, and changes in the natural level of unemployment.[19] For whatever reason, inflation and inflation expectations rose dramatically. By announcing a strict anti-inflation policy in 1979—and then following through in the 1980s—the Federal Reserve has been able to reduce inflation and inflation expectations equally dramatically. The Federal Reserve in recent years has thus reestablished a good deal of credibility by recognizing that bringing inflation down can be worthwhile in the long run despite the short-run costs of doing so. Conducting monetary policy in a way that takes account of the long-term relationship between a central bank and the private sector may therefore be the best hope for maintaining the credibility of the central bank's commitment to price stability.[20]

18 This model is due to Canzoneri, "Monetary Policy Games . . . ," and based on a model developed by Edward Green and Robert Porter in "Noncooperative Collusion Under Imperfect Price Information," *Econometrica*, January 1984, pp. 87-100.

19 For further discussion of supply shocks, money demand shocks, and changes in the natural level of unemployment in the 1970s and 1980s, see Robert J. Gordon, *Macroeconomics*. 4th edition, Little, Brown and Company, 1987, pp. 295-301; Stanley Fischer, "Monetary Policy and Performance in the U.S., Japan, and Europe, 1973-86," NBER Working Paper No. 2475, December 1987, and Stuart E. Weiner, "The Natural Rate of Unemployment . . . ," respectively.

Summary

Some economists believe that one important cause of higher inflation in the 1970s was central banks' attempt to keep unemployment at unrealistically low levels. If so, central banks' credibility in convincing the public of their commitment to price stability was tarnished. This article has explored the options available to central banks in maintaining their credibility.

The article has argued that, to the extent a credibility problem exists, solutions to the problem are themselves problematical. There are no costless ways to maintain the credibility of a central bank's commitment to price stability. Monetary growth rules remove too much discretion from a central bank operating in an environment in which financial innovation and deregulation create uncertainty about money demand and in which supply shocks can intermittently cause employment losses that perhaps should be offset by monetary policy. Nor are wage indexation and conservative central banks panaceas.

Perhaps the most promising approach is for the central bank to conduct policy in a way that takes account of the long-term nature of its relationship with the public. Even this approach has problems, however. Once inflation expectations have become imbedded in economic decisions as in the late 1970s, disinflation is likely to be accompanied by a temporary rise in unemployment until inflation expectations abate. Despite the proliferation of research analyzing the time-inconsistency and credibility problems, therefore, economists have not been able to discover a foolproof substitute for vigilance against inflation for maintaining central bank credibility.

20 Herb Taylor comes to a similar view in "Time Inconsistency: A Potential Problem for Policymakers," *Business Review*, Federal Reserve Bank of Philadelphia, March/April 1985, pp. 3-12.

Economies of Scale and Scope At Depository Financial Institutions: A Review of the Literature

By Jeffrey A. Clark

In recent years, changes in laws and regulations have greatly increased the opportunities for commercial banks and other depository financial institutions to expand their operations. Restrictions on interstate banking and intrastate branching have been liberalized in many states. In addition, limitations have been narrowed on the types of services depository institutions can offer.

While these changes have created new opportunities for individual depository institutions to grow, they have raised questions about the future structure of the banking industry. As some institutions expand and others fall prey to competitive pressures and decline or disappear, the industry's structure might come to be dominated by a small number of large diversified institutions. The market power of these institutions might allow them to keep loan rates too high and deposit rates too low, resulting in a misallocation of the nation's financial resources. The potential for resource misallocation would likely be attenuated by competitive pressures from nondepository financial institutions and from nonfmancial firms. Nevertheless, the evolving structure of the banking industry remains a source of interest and potential concern for industry observers, regulatory agencies, and policymakers.

The industry's evolving structure will depend on what types of depository institutions can remain profitable over time. Among the primary determinants of profitability will be the extent that production economies and resultant cost reductions can be achieved as firms expand their operations. If extensive cost reductions are possible, large diversified firms will potentially be more profitable than small specialized institutions.

By studying production and cost conditions that have prevailed in the past, some insight can be gained into whether the increased opportunities for growth will allow cost reduction to be achieved. This article reviews the recent literature and concludes that, in general, large diversified depository institutions have not enjoyed a large cost advantage over smaller, more specialized institutions. The article's first section discusses

Jeffrey A. Clark is associate professor of finance at Florida State University and a visiting scholar at the Federal Reserve Bank of Kansas City. The views expressed in this article are those of the author and do not necessarily represent those of the Federal Reserve Bank of Kansas City or the Federal Reserve System.

production economies and their role in influencing industry structure. The second section reviews the empirical literature on production economies at depository financial institutions. Several important issues and problems that arise in the estimation of production economies are examined in the third section. The last section summarizes the article and describes several policy implications that may be drawn from this literature.

Production economies

Two types of production economies may be achieved by individual firms in any industry—economies of scale, which are associated with firm size, and economies of scope, which relate to the joint production of two or more products.[1] Firms in an industry realize economies of scale if technology allows production costs to rise proportionately less than output when output increases. That is, economies of scale exist if per-unit or average production costs decline as output rises. Conversely, if average costs rise with output, diseconomies of scale are present. Economies of scope arise if two or more products can be jointly produced at a lower cost than is incurred in their independent production. Diseconomies of scope are present if joint production is more costly than independent production.

Industry structure is greatly influenced by the nature of production economies. If an industry's technology allows for both economies of scale and economies of scope, the industry will tend to be made up of large diversified firms.[2] These firms will be able to produce at lower per-unit costs than smaller specialized firms and can potentially use this cost advantage to gain market share. Alternatively, if technology allows neither economies of scale nor scope, small specialized firms will tend to dominate the industry. A mixture of larger diversified firms and smaller specialized firms will develop in the absence of significant economies of scale and scope.

Economies of scale

There are two kinds of economies of scale. Economies that arise from increases in the production of individual products are called product-specific economies of scale. Economies associated with increases in all of a firm's outputs are referred to as overall economies of scale.

While the two types are synonymous for a single-product firm, both types of scale economies may be present for firms that produce more than one product. For multiproduct firms, overall economies of scale occur if total costs increase proportionately less than output when there is a simultaneous and equal percentage increase in each of the firm's products. With overall economies of scale, average costs decline as the firm expands production while maintaining a constant product mix.

Product-specific economies of scale are present if a decline in the per-unit cost of producing a specific product occurs as the output of that product increases. In principle, product-specific economies of scale for each product should be measured independently from the other products in the product mix. However, in practice such a measure is not meaningful since, under joint

1 For an extensive discussion of economies of scale, see Scherer (1980).

2 In the economics literature, these institutions would be termed competitively viable. More formally, a firm is defined as competitively viable if, in the long run, no other firm can produce a given product, or product mix, at a lower per-unit cost. To an economist the concept of cost means opportunity cost. Thus, the definition of competitive viability is inclusive of all revenue and cost streams generated by alternative uses of the firm's assets. That is, if a firm's long-run costs are not at a minimum. there will be an incentive to increase profit by altering the level and/or mix of firm output.

production, it is generally impossible to change the output of one product while holding constant the output of the other **products**.[3] In spite of this problem, an approximate measure of product-specific economies of scale has been proposed and used in the empirical literature. This measure is discussed in the box on page 27.

Economies of scope

There are two types of economies of scope, global and product-specific. To define global economies of scope, it is necessary to compare the costs of both joint production and separate production, assuming a given scale for each product. For a given product mix, if the total costs from joint production of all products in the product mix are less than the sum of the costs of producing each product independently, global economies of scope are present.

Product-specific economies of scope refer to economies that arise from the joint production of a particular product with other products. If production efficiency can be enhanced by adding a particular product to a given product mix, then product-specific economies of scope exist. That is, if the cost of producing a product independently from the other products in the product mix exceeds the cost of producing it jointly, product-specific economies of scope can be realized from joint production.

Product-specific economies of scope for a given product may result from joint production efficiencies with one or a large number of products in the mix. To determine which product pairs share jointness in production, cost complementarities between all pairs of products can be computed. A cost complementarity exists between two products if the marginal cost of producing one product declines when it is produced jointly with the other.

Sources of production economies at depository institutions

The literature on the theory of the firm has hypothesized numerous ways in which economies of scale and scope might arise in production. Making better use of specialized labor and capital and spreading fixed costs over large levels of output are usually cited as the predominant sources of economies of scale. Most economies of scope are thought to arise from the joint usage of a fixed resource.

Consistent with the theory of the firm, research on production by depository institutions often points to these important sources of both economies of scale and scope: specialized labor, computer and telecommunications technology, and information. For example, at small depository institutions, labor is unlikely to perform specialized functions. Tellers and loan officers probably process a variety of loan and deposit accounts since they are likely to be underutilized in handling specialized products. Their unspecialized labor is then a fixed input that can be shared in the production of a number of products, with the potential to create economies of scope. As these smaller institutions grow, they may be able to fully employ more specialized labor in producing some or all of their products. If the expertise of specialized tellers and loan officers results in the processing of a greater volume of deposit and loan accounts per unit of labor, then per-unit labor costs can be reduced through increased specialization. In this example, increased size may result in production efficiencies through the substitution of economies of scale for economies of scope.

The adoption of computer and telecommunica-

[3] **For expanded discussion of this problem, see Fuss and Waverman (1981).**

tions equipment can provide another basis for both economies of scale and scope at depository institutions. Despite the large set-up costs required, computer and other electronic funds transfer equipment can process a large volume of transactions at a small additional cost per transaction. As depository institutions increase the number of transactions of all types that can be performed by this equipment, it may be possible to reduce the per-unit cost of the firm as a whole as well as for individual products. Embracing this technology may provide a basis for both overall and product-specific economies of scale. In addition, any excess capacity of the equipment could be used to process other types of accounts at a small additional cost per transaction, thus realizing economies of scope.

Economies of scale and scope may also accompany information production. Before lending decisions can be made, credit information must be gathered and analyzed. Once gathered, however, this information can be reused in other lending decisions. Where the cost of reusing information is less than the independent cost of its production, reuse can help reduce the incremental costs of extending additional credit. If the information is reused to make similar loans to the same customer or to other customers in the same region or industry, it will provide a source of economies of scale. Alternatively, if the information can be used to make unrelated types of loans to the institution's customers, it may serve as a source of economies of scope.

A review of the empirical literature

Most of the evidence about the existence and extent of production economies at depository institutions comes from the empirical estimation of statistical cost functions. In developing these functions, researchers begin with the microeconomic principle that production costs depend on input prices and the level and composition of output.[4] After defining these variables, the researcher selects a statistical function to explicitly relate production costs to outputs and input prices. The most frequently selected statistical function is the transcendental logarithmic or translog function. This function is usually selected because it is flexible enough to yield both economies and diseconomies of scale at different output levels and to provide information on scope economies by incorporating interdependencies between products.[5]

Once the statistical function is selected and specified, the researcher estimates the parameters of the function using sample data. The estimated parameters and sample data are then used to construct empirical measures of the various types of scale and scope economies discussed in the previous section. A discussion of the most frequently used empirical measures is presented in the box on page 27. Technical statements of each measure are presented in Appendix B.

Empirical evidence

The 13 studies reviewed in this article anempted to estimate economies of scale and scope for credit unions, savings and loan associations, or commercial banks. Each study used a translog statistical cost function and employed similar measures of economies of scale and scope. The studies' results suggest four broad conclusions: First, overall economies of scale appear to exist

[4] This functional relationship follows from the property of duality between the production and cost functions. When the statistical cost function is being estimated with cross-sectional data, it may be necessary to include other variables that may induce interfirm variation in cost. Among the variables most commonly included are the number of branches and affiliation with a holding company.

[5] An example of the general form taken by the translog function appears in Appendix B.

only at low levels of output with diseconomies of scale at large output levels. Second, there is no consistent evidence of global economies of scope. Third, there is some evidence of cost complementarities (product-specific economies of scope) in production. Finally, these results appear to be generally robust across the three types of institutions, as well as across different data sets and product and cost definitions.

Twelve of the 13 studies report significant overall economies of scale at relatively low levels of output (Table 1, column 2). Only Mester fails to find any evidence of scale economies, and then only for savings and loan associations below $100 million of deposits.[6] Only two studies, however, find significant overall economies of scale above $100 million of deposits (Table 1, column 3). Moreover, the authors of one of these—Goldstein, McNulty, and Verbrugge—do not directly control for potential scope economies, and the authors of the other study—Benston, Hanweck, and Humphrey—repon scale economies only for large branch banking organizations.[7] Several authors report greater economies of scale among branch banking institutions, but when an augmented measure of overall economies of scale is employed to control for the interdependency between the number of offices and the number of accounts serviced, the cost advantage of branch banks seems to disappear.[8]

As already noted, it is not conceptually possible to measure product-specific economies of scale without ambiguities, so it may not be surprising that only four of the 13 studies report evidence on this type of production economy.[9] The results presented in these four studies do not support a conclusion of widespread product-specific economies of scale (Table 1, column 4). Both H.Y. Kim and Mester report product specific economies of scale for mortgage loans. However, H.Y. Kim and Gilligan, Smirlock and Marshall also report product specific diseconomies of scale for several products.[10] Benston, Berger, Hanweck, and Humphrey report estimates of the marginal cost of production for five products by size class. However, they acknowledge that the negative marginal costs reported for some products are "implausible" and most likely indicate some type of estimation problem.[11]

Eleven of the studies compute a measure of global economies of scope. However, only three report evidence of statistical significance for their measure. Further, in two of the three studies that report statistically significant global economies of scope, the statistical cost function that was estimated contained only two broadly defined products.[12] Only M. Kim reported statistically

6 See Mester (1987).

7 See Goldstein, McNulty, and Verbrugge (1987); and Benston, Hanweck, and Humphrey (1982).

8 See Appendix B for a presentation of the augmented measure of overall economies of scale used to control for the relationship between the number of offices and the number of accounts.

9 Appendix B presents several methods proposed by these authors for measuring product-specific economies of scale.

10 These products include nonmortgage loans, investment services, total loans and total deposits.

11 The authors suggest that the most likely estimation problems are the presence of multicollinearity and the loss of degrees of freedom, both resulting from the large number of parameters that must be estimated when the translog function is used. See Benston et al. (1983).

12 Gilligan, Smirlock, and Marshall (1984) include total deposit accounts and total loan accounts as the only two products in the cost function they estimate. Gilligan and Srnirlock (1984) estimate two statistical cost functions, each with a pair of products. The product pairs employed in the two cost functions are, respectively, the total dollar amounts of demand and time deposits, and the total dollar amounts of total loans outstanding and total securities held.

TABLE 1

Summary of results of studies reviewed

Authors	Significant economies of scale: Overall: Below $100 million in deposits	Significant economies of scale: Overall: Above $100 million in deposits	Significant economies of scale: Product specific	Significant economies of scope: Global	Significant economies of scope: Cost complementarities
Murray and White (1983)	yes	no	no measure-	no	yes
H.Y. Kim (1987)	yes[c]	no	yes[o,j] no[p,j]	no[k]	no[k]
Mester (1987)	no	no	yes[o]	no	no
Goldstein, McNulty, & Verbrugge (1987)	yes[a]	yes[a]	no measure	no measure	no measure
LaCompte and Smith (1986)	yes[b]	no	no measure	no	yes(1978) no(1983)
Benston, Hanweck, & Humphrey (1982)	yes[d,f] no[e]	yes[d,f] no[e]	no measure	no measure[l]	no measure[l]
Benston, Berger, Hanweck, & Humphrey (1983)	yes	no	yes[r,j]	no	yes[j]
Gilligan and Smirlock (1984)	yes	no	no measure	yes[m]	no measure
Gilligan, Smirlock, & Marshall (1984)	yes	no	no[s,j]	yes[m]	no measure
M. Kim (1986)	yes	no	no measure	yes[n] no	yes[n] no
Lawrence and Shay (1986)	no[g]	no[h]	no measure	no[o]	yes[o]
Berger, Hanweck, & Humphrey (1987)	yes	no	no measure	no	yes
Kolari and Zardhooki (1987)	no[e] yes[d]	no[e] no[d,i]	no measure	no	yes

Notes:
- a: Did not control for economies of scope.
- b: Up to $50 million in total deposits.
- c: Reports diseconomies of scale to nonmortgage lending.
- d: Lknotes branch banking.
- e: Denotes unit banking.
- f: Reports diseconomies of scale if an augmented global scale economies measure is utilized.
- g: Reports increasing returns to scale in 1980 and 1981 only.
- h: No diseconomies of scale found in the upper two quartiles as high as 52.5 billion in 1980 and 1981.
- i: Up to $100 million in total deposits.
- j: Provides no statistical tests.
- k: Reports scope economies but without tests of statistical significance.
- l: Employed Divisia Index for output.
- m: Test of nonjointness restrictions used only one pair of outputs.
- n: Denotes a no-aggregation model.
- o: Diseconomies of scope found between loans and investments.
- p: For mortgage loans only.
- q: Reports diseconomies for nonmortgage lending and investment services.
- r: Reports computed marginal costs for selected products and arbitrarily chosen deposit size classification.
- Reports diseconomies of scale for total loans and total deposits.

significant global economies of scope for a more disaggregated product mix.[13] The last two columns of Table 1 summarize the estimates of global and product-specific (cost complementarities from joint production) economies of scope.

Although the empirical evidence does not support a conclusion of global economies of scope from joint production, many of the studies report some evidence of cost complementarities between pairs of products. When the translog function is estimated, evidence of a cost complementarity between any two products is given by a negative and statistically significant parameter estimate on the cross-product term between the two products. Table 2 lists all product pairs for which the estimated cross-product term is statistically significant. Inspection of this table indicates that some evidence of cost complementarities can be found in a number of studies and among a variety of different product pairs. The strongest evidence of cost complementarities occurs in the joint production of two product pairs: total loans and total deposits, and investments and mortgage loans.[14] However, diseconomies of joint production were also reported between two related product pairs: investments and total loans, and total loans and total deposits for branch banks with total deposits below $100 million.[15]

Issues and problems

Several issues and problems may have influenced the results discussed in the preceding section. These issues and problems are both conceptual and methodological in nature. The problems tend to limit, but not eliminate, the usefulness of the empirical conclusions in drawing policy implications.

Defining bank costs and output

The banking literature is divided over the conceptual issue of the appropriate definition of bank output, and consequently on the related issue of defining bank costs. In general, researchers take one of two approaches.[16] These alternative approaches are labeled the "intermediation approach" and the "production approach."[17] No consensus has developed favoring one of the definitions over the other, and reasonable arguments have been made for both approaches.

Under the intermediation approach, depository financial institutions are viewed as producers of services related directly to their role as an intermediator in financial markets. That is, they are viewed as collecting deposits and purchasing funds to be subsequently intermediated into loans and other assets. In this approach, deposits are treated as inputs along with capital and labor. Those authors who adopt this approach generally define the institution's various dollar volumes of earning assets as measures of output. Also con-

13 In his study of Israeli banks, Kim (1986) defined several alternative product mixes as combinations of four distinct products: demand deposits, foreign currency, loans, and securities. His results indicate that global economies of scope only occur when the four products appear separately in the cost function. Kim reports an absence of global economies of scope for all other combinations of these four products.

14 A cost complementarity between total loans and total deposits is reported in Berger, Hanweck, and Humphrey (1987); Gilligan, Smirlock, and Marshall (1984); and Lawrence and Shay (1986). A cost complementarity between investments and mortgage loans is reported in LaCompte and Smith (1986), and Mester (1987).

15 The diseconomy of the first type is reported in Lawrence and Shay (1986). The second type of diseconomy is reported in Berger, Hanweck, and Humphrey (1987). Lawrence and Shay report an additional diseconomy of joint production between nonbank activities and total deposits.

16 The approaches taken in the 13 papers reviewed here appear in the second column of Appendix A.

17 Discussions of these two approaches can be found in a number of recent papers including Humphrey (1987); Mester (1987a); and Berger, Hanweck, and Humphrey (1987).

TABLE 2
Significant cost complementarities

Output Pairs	Author(~)	Year(~)	Sign
Consumer and mortgage loans	LaCompte and Smith	1978	negative
Investments and total loans	Gilligan and Smirlock	1973-78	negative
	Lawrence and Shay	1982	positive
Nonbank activity and total loans	Lawrence and Shay	1982	negative
Total deposits and total loans	Lawrence and Shay	1982	negative
	Gilligan, Smirlock, & Marshall	1978	negative
	Berger, Hanweck, & Humphrey	1983	negative*
Investments and mortgage loans	Mester	1982	negative
	LaCompte and Smith	1978	negative
Nonbank activity and investments	Lawrence and Shay	1982	negative
Total deposits and investments	Lawrence and Shay	1982	negative
Nonbank activity and total deposits	Lawrence and Shay	1982	positive
Time deposits and demand deposits	Gilligan and Smirlock	1973-78	negative

* Negative for branch banks with deposits > $100 million in total deposits; positive for branch banks < $100 million in total deposits.

sistent with this approach, costs are defined to include both interest expense and total costs of production.

The production approach, on the other hand, views depository institutions as producers of services associated with individual loan and deposit accounts. These account services are produced using capital and labor. Under this approach, it follows that the number of accounts of each type are the appropriate definitions of outputs. Total costs are defined exclusive of interest costs.

Conceptually, the intermediation and production approaches are very different. In reviewing the literature, it is surprising that the empirical results do not appear to be sensitive to the approach taken in defining outputs and costs. Why this should be the case is unclear. However, one possibility is that other issues, as discussed below, are more important.

Data

One of two types of data has been employed in nearly all recent attempts at estimating statistical cost functions for depository institutions. The data are drawn either from Call Report and financial statement data (as reported to the Federal Deposit Insurance Corporation, the Federal Savings and Loan Insurance Corporation, and the National Credit Union Share Insurance Fund), or from the Functional Cost Analysis (FCA) program conducted by the Federal Reserve System.

Each of these two sources of data offers advantages and disadvantages. An advantage of the FCA data is that they are constructed using simple cost accounting techniques to allocate costs among several distinguishable banking functions. In addition, these data include information on the number and average size of a variety of deposit and loan products. However, the generalization of the results obtained using FCA data to all depository institutions may be inappropriate for several reasons. Because the FCA program is voluntary, subscribing banks might be either high-cost institutions interested in identifying areas for cost reduction or low-cost firms that place greater emphasis on cost control. Further, the FCA data are heavily skewed toward small banks.[18] Finally, the procedures used to allocate costs are sometimes imprecise and may induce unknown bias in parameter estimates when the FCA data are used to estimate statistical cost functions.[19]

An advantage of Call Report and financial statement data is that they provide information on a much wider range of institutional size and impose uniform reporting requirements. The empirical results obtained using these data, therefore, should be more generally applicable. However, this source of data also imposes limitations. First, the absence of information on numbers of deposit and loan accounts and average account size make this source of data unsuitable for use under the production approach. Further, there is some evidence that the average account size and institution size are positively correlated. Thus, a failure to control for average account size under the intermediation approach may tend to overstate any finding of economies of scale. Second, data on some banking functions such as loan commitments, standby letters of credit, safety deposit and trust activity have only recently, if at all, been reported in these data. Finally, it is questionable whether financial statement data can be used to construct meaningful proxies for the input prices, given the high level of aggregation at which these data are reported.

---

18 As of 1986, only 490 banks participated in the program. Of this number, 416 were under $200 million in total deposits.

19 In some cases, the allocations are made according to the judgment of the participating banker (e.g., wages and salaries). In other instances, the allocations are performed by computer algorithms developed for a representative bank using "experience factors" that are derived from previous data. For additional discussion of the allocation rules, see the *Introduction to Functional Cost Analysis: 1986 Average Banks.*

Level of aggregation and limitations of the translog functional form

Two closely related issues arise in the estimation of scale and scope economies: the appropriate level of aggregation and the suitability of the translog functional form for use with data from depository institutions. Theoretically, a measure of each distinct product offered by depository institutions should be included in the estimated function. However, the feasibility of doing this is usually limited by the availability of data and the use of a translog functional form. The larger the number of distinct products that are defined, the greater the likelihood that institutions included in the sample do not produce some of the products. Since the translog function expresses each input price and the output of each product in logarithmic form, the values of these variables must be strictly greater than zero. If a high level of disaggregation is chosen to increase the ability to identify jointness in production, then smaller and more specialized depository institutions will need to be deleted from the sample. Alternatively, if the level of aggregation in defining products is high enough to provide positive values for the output of all defined products for all institutions in the sample, then much of the information on efficiencies from joint production may be lost.[20]

A second problem involving the level of disaggregation and the translog functional form arises in attempting to compute measures of product specific economies of scale and global economies of scope. The computation of these measures requires the assumption of a zero level of output for at least one of the products being produced. However, the translog cost function will always yield zero total costs whenever the output of even one product is zero. To circumvent this problem, most researchers compute total costs by choosing an arbitrarily small but nonzero value for use in place of zero. This procedure has two drawbacks. First, the arbitrarily chosen value is usually well outside the bounds of the data. As a consequence, the confidence intervals around any computed values for these measures will be extremely wide. Second, the conventional measure of global economies of scope can be made to yield scope diseconomies. This result can be insured by replacing all zero outputs with a sufficiently small nonzero value.[21]

A third source of problems involving the level of disaggregation and the translog function arises from the number of parameters that must be estimated. As more products are defined and included in the statistical cost function, the number of parameters that must be estimated increases disproportionately.[22] For depository institutions, the number of products that must be defined to yield any meaningful level of disaggregation is large. With the necessity of including linear, quadratic, and cross-product terms for all defined products and input prices, the likelihood of severe multicollinearity would appear to be high. In this case, it may not be possible to identify individual parameter estimates. Any statistical tests will be imprecise since the standard errors of the parameter estimates are likely to be large.[23]

20 Kim (1986) reports evidence that suggests if product definitions are drawn too broadly, the resulting parameter estimates will be biased against the identification of significant economies of scope.

21 A thorough discussion of this problem can be found in Benston et al. (1983).

22 Mester (1987a) has noted that the addition of one input and one product to a translog function consisting of three inputs and three products increases the number of parameters that must be estimated from 28 to 45.

23 The author of this article estimated a translog cost function with seven defined products using a sample of 190 commercial banks in the Denver SMSA in 1987. All of the included banks had nonzero values for each defined product. The products

Other incentives for joint production

The concept of cost in economics is synonymous with opportunity cost, not accounting cost. Thus, in principle, the measurement of economies of scale and scope using a statistical cost function should attempt to measure the total costs of production in terms of opportunity costs rather than accounting costs. While technology may provide opportunities for the sharing of inputs, the decision to add product lines will depend ultimately on whether the additional product will increase after-tax, risk-adjusted returns. The focus on accounting costs results in the exclusion of any revenue and tax-related incentives for adding product lines—such as a reduction in earnings volatility from increased diversification—that are not rooted in production efficiencies and may even increase per-unit accounting costs.[24]

Summary

Care should be exercised in attempting to use the existing empirical literature as a sole basis for policy. At present, no systematic attempts have been made at conducting a sensitivity analysis of the empirical results to the issues and problems discussed above. Further, it is difficult to assess the severity of these problems by examining the existing literature because differences among studies are sufficiently large to prevent drawing conclusions on specific issues.

Finally, the studies reviewed in this article predate the granting of new securities, insurance, mutual funds, and other powers to depository institutions and therefore cannot be used to draw inferences about their likely impact on costs. This is particularly true since the size of any impact will depend importantly upon whether the new powers are granted directly to institutions or can be offered only through affiliates of bank holding companies.

Conclusions and policy implications

A review of the empirical evidence presented in 13 separate studies of economies of joint production for depository institutions yields several tentative results. First, the empirical evidence appears to support a conclusion of significant overall economies of scale only for depository institutions of relatively small size—less than $100 million in total deposits. Second, the empirical evidence does not appear to support a conclusion of global economies of scope. Third, there appears to be some evidence of economies in joint production among specific pairs of products that might be offered by depository institutions.

The three results listed above suggest several tentative policy conclusions. Taken together, the evidence implies that the smallest, most specialized of depository institutions may be at a cost disadvantage relative to larger, more diversified

---

appearing in the cost function included the dollar value of transactions deposits, time deposits, investments, real estate loans, installment loans, credit card loans, and commercial loans, respectively. Under the assumption that at the margin all banks in the market faced the same input prices, use of the translog cost function required the estimation of 36 parameters. The estimation of the translog cost function produced an adjusted R^2 of .9783 and an F-statistic of 245.782. However, of the 36 parameters only four were statistically significant at the 5 percent level of significance. The variance decomposition collinearity diagnostics provided in SAS produced "high" condition indices for all but five of the variables. Further, there were numerous instances in which variables with high condition indices contributed strongly (exhibited a variable proportion greater than 0.5) to the variances of two or more variables. These results are indicative of a severe multicollinearity problem.

24 Other incentives may include greater use of off-balance sheet activities to avoid regulatory taxes imposed by risk-based capital requirements and deposit insurance premiums, and joint customer demand for banking services that arise from a desire to reduce transactions costs. See Baer and Pavel (1988) for a recent analysis of the regulatory tax imposed by minimum capital requirements and deposit insurance premiums.

institutions. These smaller institutions are likely to be faced with the necessity of increasing both the scale and scope of their operations to remain cost competitive. Failure to achieve sufficient growth and to exploit available cost complementarities may drive these depository institutions from the market or cause them to be absorbed by other more cost-efficient institutions. However, the evidence also suggests that once overall scale economies have been exhausted, there will still be opportunities for the smaller, less diversified depository institutions. The absence of strong global economies of scope, combined with evidence of several cost complementarities, will probably provide a number of market niches for these smaller institutions.

From a policy perspective, the absence of a cost advantage for the largest, most diversified depository institutions appears to minimize any concern that the banking industry will be dominated by a few large depository financial institutions. The lifting of restrictions on interstate banking and intrastate branching might help consolidate resources in states that have prohibited or severely limited branch banking by permitting small banks to achieve a more efficient scale of production. The absence of significant scope economies suggests, however, that the lifting of these restrictions is unlikely to require significant adjustment in product mix.

In light of the issues and problems raised in this article, there is ample room for more research. Future efforts should address questions like these: Is there a better statistical function for use in measuring economies of scale and scope than the translog cost function? What is the appropriate level for the disaggregation of output for depository institutions? What is the best way to broaden the focus to include incentives for joint production? And, as new powers are granted to depository institutions, how will this affect their production efficiencies?

Empirical Measures of Production Economies

Researchers have developed empirical measures for both economies of scale and economies of scope. Overall economies of scale are typically measured by computing the sum of the output cost elasticities of individual products. The output cost elasticity for a product is the percentage change in production costs that occurs for a given percentage change in the output of the product. And, the sum of the individual output cost elasticities is equivalent to the percentage change in costs that results from an equal percentage change in the output of all products. When this measure of overall economies of scale is equal to one at a given level of overall output, there are constant returns to scale. Thus, no additional production efficiencies can be achieved in this range of production. If this measure of overall scale economies is significantly less than one, then there are increasing returns to scale and production efficiencies will be realized in this range of production. Conversely, if this measure is significantly greater than one, there are decreasing returns to scale and production inefficiencies will be realized.

While product-specific economies of scale cannot be measured without ambiguities, an approximate measure has been proposed and utilized in several cost studies. This measure makes use of the theoretical relationship between the marginal cost, average cost, and economies of scale. Where the marginal cost of producing a product is less than average cost at a given level of output, average cost is declining in that range of output, implying economies of scale. Conversely, when marginal cost is greater than average cost, average cost is increasing, implying diseconomies of scale. To approximate this relationship in a multiproduct setting, a new cost concept labeled "average incremental cost" (AIC) is utilized. AIC is defined as the addition to total cost of producing a specific level of a product as opposed to not producing it at all, divided by the level of output of the product. Then the AIC can be expressed as a ratio to the marginal cost of producing this level of output. If this ratio is greater than one, this is viewed as evidence of product-specific economies of scale for the range of output levels between zero and the level at which AIC and MC are evaluated, since it implies that average costs are declining. If the ratio is less than one, product-specific diseconomies of scale is implied.

Global economies of scope are measured by computing the cost differential that would arise between the independent and joint production of specific output levels of all products. This cost differential is then generally scaled by dividing by the total costs of joint production. This measure will have a value greater than zero when there are global economies of scope, and a negative value when diseconomies are present.

As an alternative to computing the preceding measure, researchers have demonstrated that a sufficient condition for global economies of scope is the existence of cost complementarities among all pairs of products in the product mix. A cost complementarity occurs when the marginal cost of producing one product declines with an increase in the level of production of another.

Product-specific economies of scope are measured in several alternative ways. One common measure is to compute the cost increase or decrease that arises from producing a specific product both independently from, and jointly with, the remaining product mix and expressing it as a percentage of the costs of joint production. If this ratio is greater than one, product-specific economies of scope are implied. If the ratio is less than one, diseconomies of scope exist.

Other alternative ways of identifying a cost complementarity between any two products in the product mix involve an assessment of how joint production of two products affects the marginal cost of producing each product. When parameter estimates from a translog statistical cost function are used, it can be shown that a necessary condition for the marginal cost of producing a product to decline with an increase in the production of a second product, referred to here as a pairwise cost complementarity, requires their cross-product term to be negative and statistically different from zero. However, while a negative cross-product term is consistent with the existence of a cost complementarity, it is not sufficient. Any reduction in marginal costs from the joint production of the two products may be offset by rapidly rising marginal costs from one or both of the two products. When the translog function is estimated, it can be shown that a sufficient condition for a cost complementarity between two products requires that the cross-product term not only be negative but also greater in absolute value than the product of the output elasticities of the two products being considered. A statistical test of this condition (test of nonjointness) is carried out by testing the parameter restrictions that would be required for nonjointness in the production of the two products.

Appendix A
Summary of Studies Reviewed

Authors	Approach	Data	Outputs	ECSCA*	ECSCO*
Murray and White (1983)	Intermediation	61 Canadian Credit Unions (1976-77)	y_1,y_2,y_{14}	OSA(1)	PSSO(2)
H. Y. Kim (1986)	Intermediation	61 Canadian Credit Unions (1976-77)	y_1,y_2,y_{14}	OSA(1) PSSA(1)	GSO(1) GSO(2) PSSO(4)
Mester (1987)	Intermediation	149 Calif. S&Ls (1982)	y_1,y_3,y_{14}	OSA(1)	GSO(1) PSSO(1)
Goldstein, McNulty, & Verbrugge (1987)	Production	FSLIC Insured S&Ls (1978-81)	y_{13}	OSA(1)** OSA(2)**	No measure
LaCompte & Smith (1986)	Intermediation	S&Ls Ninth Dist. FHLBB (1978-83)	y_1,y_2,y_3	OSA(1)	PSSO(2)
Benston, Hanweck & Humphrey (1982)	Production and Intermediation	FCA Data (1975-78)	Divisia Index	OSA(1)**	No measure
Benston, Berger, Hanweck & Humphrey (1983)	Production	FCA Data [deposits less than one billion] (1978)	$y_4,y_5,y_6,$ y_7,y_8	OSA(1) OSA(2)	PSSO(3) PSSO(4)
Gilligan & Smirlock (1984)	Production	Financial Statement Data, 2700 banks (1973-78)	$y_3,y_4,y_5,$ y_9	OSA(1)	PSSO(3)
Gilligan, Smirlock, & Marshall (1984)	Production	FCA Data (1978)	y_9,y_{10}	PSSA(2)	PSSO(3)
M. Kim (1986)	Intermediation	17 Israeli Banks (1979-82)	$y_3,y_4,y_9,$ y,,	OSA(1) OSA(2)	GSO(1) PSSO(1)
Lawrence & Shay (1986)	Intermediation	FCA Data (1979-82)	y_3,y_4,y_9 y_{12}	OSA(1)	PSSO(4)
Berger, Hanweck & Humphrey (1987)	Production	FCA Data (1983)	y_4,y_5,y_6 y_7,y_8	OSA(1) EPSA	GSO(1) EPSUB
Kolari & Zardhooki (1987)	Production	FCA Data (1979-1983)	y_3,y_4,y_5 y_9,y_{10}	OSA(1)	GSO(1) PSSO(2)

Notes: *See Appendices A and B for definitions of the abbreviations for the measures of economies of scale and scope employed in this table.

** indicates the use of a Divisia Index for output. Other outputs are denoted as follows: y, =mortgage loans; y_2=consumer loans; y, = investments; y, =demand deposits; y_5=time deposits; y_6=real estate loans; y, =commercial loans; y_8= installment loans; y_9=total loans; y_{10}=total deposits; y,, =foreign currency; y_{12}=nonbank activities; y,,=total assets; and y,,=other loans.

Appendix B
Empirical Measures of Economies of Scale and Scope

I. TRANSLOG STATISTICAL COST FUNCTION

$$lnTC = B_o + \sum_i B_i lny_i + \sum_k C_k lnp_k + (1/2)\sum_i \sum_j D_{ij} lny_i lny_j;$$

$$+ (1/2)\sum_k \sum_l E_{kl} lnp_k lnp_l + \sum_i \sum_k F_{ik} lny_i lnp_k + e,$$

where ln denotes the logarithm; $y_i (i=1,\ldots,m)$ denotes the ith output; $p_k (k=1,\ldots,n)$ denotes the kth input price; B_o, B_i, C_k, D_{ij}, E_{kl}, F_{ik} are the parameters to be estimated and e represents the random error term.

II. OVERALL ECONOMIES OF SCALE

A. Overall or Plant Economies of Scale

$$OSA(1) = \sum_i \frac{\partial lnTC}{\partial lny_i} = \sum_i \epsilon_i, \text{ where } \epsilon_i$$

is the output cost elasticity for product i. $OSA(1) < 1$ indicates overall economies of scale. $OSA(1) > 1$ indicates overall diseconomies of scale.

B. Augmented or Firm Economies of Scale

$$OSA(2) = \sum_i \frac{\partial lnTC}{\partial lny_i} + \frac{\partial lnTC}{\partial lnB} \times \frac{\partial lnB}{\partial lny_i},$$

where B is the number of branches operated by the depository institution. $OSA(2) < 1$ indicates overall economies of scale. $OSA(2) > 1$ indicates overall diseconomies of scale.

III. PRODUCT-SPECIFIC ECONOMIES OF SCALE

A. Average Incremental Costs

$PSSA(1) = [(IC_i/TC)/\epsilon_i]$, where $\epsilon_i = \frac{\partial lnTC}{\partial lny_i}$

$TC = C(y_i,\ldots,y_m)$ and $IC_i = [C(y_i,\ldots,y_m) - C(y_1,\ldots,y_{i-1}, 0, y_{i+1},\ldots,y_m)]$. $PSSA(1) > 0$ indicates product-specific economies of scale for product y_i. $PSSA(1) < 0$ indicates product-specific diseconomies of scale for product y_i.

B. Declining Marginal Cost

$$PSSA(2)=\frac{\partial^2 TC}{\partial y_i^2}=(\frac{TC}{y_i^2})[\frac{\partial lnTC}{\partial lny_i^2}+(\frac{\partial lnTC}{\partial lny_i})(\frac{\partial lnTC}{\partial lny_i}-1)].$$

If $PSSA(2) < 0$ then marginal costs of product y, are declining. This implies product-specific economies of scale for product y_i. $PSSA(2) > 0$ implies increasing marginal costs and product-specific diseconomies of scale for product y_i.

IV. EXPANSION PATH SCALE ECONOMIES

$$EPSA=\{\sum_i[((y_i^B-y_i^A)/y_i^B)(C(y_i^B)-C(y_i^A)/C(y_i^B))]\times \partial lnTC^B\},$$

where y_i denotes the level of output of product i produced by small Firm A or large Firm B. C( ) denotes the total cost of producing level y_i of product i by each type of firm. If $EPSA < 1$ this implies economies of scale along an expansion path including firms A and B. If $EPSA > 1$ this implies diseconomies of scale along this expansion path.

V. GLOBAL ECONOMIES OF SCOPE

A. Global Economies of Scope

$$GSO(1)=\{[C(y_1,o,\dots,o)+\dots+C(o,\dots,o,y_m)]-C(y_1,\dots,y_m)\}/C(y_1,\dots,y_m),$$

where C( ) denotes the total costs of production. If $GSO(1) > 0$ then there are global economies of scope. If $GSO(1) < 0$ there are global diseconomies of scope.

B. Disjoint-Group Economies of Scope

$$GSO(2)=\{[C(y_1,\dots,y_j)+C(y_{j+1},\dots,y_m)]-C(y_1,\dots,y_m)\}/C(y_1,\dots,y_m), \text{ where}$$

C( ) denotes the total costs of production. $GSO(2) > 0$ denotes economies of scope in production. $GSO(2) < 0$ denotes diseconomies of scope.

VI. PRODUCT-SPECIFIC ECONOMIES OF SCOPE

A. Product-Specific Economies of Scope

$$PSSO(1)=\{[C(y_1,\dots,y_{i-1},o,y_{i+1},\dots,y_m)+C(o,\dots,o,y_1,o,\dots,o)]-C(y_1,\dots,y_m)\}/C(y_1,\dots,y_m)$$

where C( ) denotes the total costs of production. $PSSO(1) > 0$ implies product-specific economies of scope. $PSSO(1) < 0$ implies product-specific diseconomies of scope.

B. Cost Complementarities

$$PSSO(2) = \frac{\partial^2 TC}{\partial y_i \partial y_k} = \left(\frac{TC}{y_i y_k}\right)\left[\frac{\partial^2 lnTC}{\partial lny_i \partial lny_k} + \left(\frac{\partial lnTC}{\partial lny_i}\right)\left(\frac{\partial lnTC}{\partial lny_k}\right)\right]$$

$PSSO(2) < 0$ implies that an increase in the level of production of product y_k reduces the marginal cost of producing product y_i. Thus $PSSO(2) < 0$ implies product-specific economies of scope between products y_i and y_k. Conversely, $PSSO(2) > 0$ implies product-specific diseconomies of scope between products y_i and y_k.

, C. Test of Nonjointness

From $PSSO(2)$, nonjointness implies $(\partial^2 TC/\partial y_i \partial y_k) = 0$. At any nonzero level of production of y_i and y_k, $(TC/y_i y_k) > 0$. Therefore, nonjointness requires

$$PSSA(3) = \left[\frac{\partial lnTC}{\partial lny_i \partial lny_k} + \left(\frac{\partial lnTC}{\partial lny_i}\right)\left(\frac{\partial lnTC}{\partial lny_k}\right)\right]^5 0.$$

From the translog this implies the restrictions that

$$[D_{ij} + \epsilon_i \times \epsilon_k] = 0, \text{ where}$$

$$\epsilon_i = \frac{\partial lnTC}{\partial lny_i} = B_i + \sum_j D_{ij} lny_j + \sum_k F_{ik} lnp_k.$$

The parameter restrictions can be imposed and a likelihood ratio test of the restrictions can be conducted.

D. Pairwise Cost Complementarities

A necessary condition for $(\partial^2 TC/\partial y_i \partial y_k) < 0$, is that the value of $(\partial lnTC/\partial lny_i \partial lny_k) < 0$. This follows because, as in $PSSO(3)$, $(TC/y_i y_k) > 0$. Further, from theory, $MC_i = (\partial TC/\partial y_i) > 0$, so that $(\partial lnTC/\partial lny_i) = (\partial TC/\partial y_i)(y_i/TC) > 0$. Therefore, a necessary condition for the existence of a cost complementarity between products y_i and y_k, when estimating the translog cost function, is

$$PSSO(4) = \frac{\partial^2 lnTC}{\partial lny_i \partial lny_k} = D_{ik} < 0.$$

VII. EXPANSION PATH SUBADDITIVITY

$$EPSUB = \{[C(Y^A) + C(Y^D) - C(Y^B)]/C(Y^B)\},$$

where $Y^A = (y_1^A, y_2^A, \ldots, y_m^A)$ is the product-mix of small firm A, $Y^B = (y_1^B, y_2^B, \ldots, y_m^B)$ is the product-mix of large firm B, and $Y^D = (Y^B - Y^A)$; $\$ \geq \$ \geq 0 \; \forall i$. $EPSUB > 0$ implies a cost advantage for large firm A.

References

Baer, Herbert L. and Christie A. Pavel. "Does Deregulation Drive Innovation?" *Economic Perspectives.* Federal Reserve Bank of Chicago, March/April 1988, 3-16.

Bailey, Elizabeth E. and Ann F. Friedlaender. "Market Structure and Multiproduct Industries." *Journal of Economic Literature.* September 20, 1982, 1024-1048.

Baumol, William J., John C. Panzar, and Robert D. Willig. *Contestable Markets and Industry Structure.* New York, Harcourt Brace Jovanovich, Inc., 1982.

Benston, George J., Gerald A. Hanweck, and David Humphrey. "Scale Economies in Banking." *Journal of Money. Credit. and Banking,* 14, Part 1, November 1982, 435-456.

Benston, George J., Allen N. Berger, Gerald A. Hanweck, and David Humphrey. "Economies of Scale and Scope." *Proceedings of a Conference on Bank Structure and Competition.* Chicago, Federal Reserve Bank of Chicago, 1983.

Berger, Allen N., Gerald A. Hanweck, and David B. Humphrey. "Competitive Viability in Banking." *Journal of Monetary Economics,* 20, 1987, 501-520.

Functional Cosr Analysis: 1986 Average Banks. Federal Reserve Banks, 1986.

Gilligan, Thomas W. and Michael L. Smirlock. "An Empirical Study of Joint Production and Scale Economies in Commercial Banking." *Journal of Banking and Finance,* 8, 1984, 67-77.

Gilligan, Thomas W., Michael Smirlock, and William Marshall. "Scale and Scope Economies in the Multi-Product Banking Firm." *Journal of Monetary Economics.* 13, 1984, 393-405.

Goldstein, Steven J., James E. McNulty, and James Verbrugge. "Scale Economies in the Savings and Loan Industry Before Diversification." *Journal of Economics and Business,* 1987, 199-207.

Humphrey, David B. "Cost Dispersion and the Measurement of Economies in Banking." *Economic Review,* Federal Reserve Bank of Richmond, May/June 1987, 24-38.

Kim, H. Youn. "Economies of Scale and Scope in Multiproduct Financial Institutions: Further Evidence from Credit Unions." *Journal of Money, Credit, and Banking,* 18. May 1986, 220-226.

Kim, Moshe. "Banking Technology and the Existence of a Consistent Output Aggregate." *Journal of Monetary Economics,* 18. 1986, 181-195.

Kolari, James and Asghar Zardhooki. *Bank Cosr. Structure, and Performance.* Lexington, Mass., D.C. Heath, 1987.

LaCompte, Richard L.B. and Stephen D. Smith. "The Impact of Regulation on Cost Structures: The Case of the Savings and Loan Industry." Unpublished manuscript, September 1986.

Lawrence, Colin and Robert Shay. "Technology and Financial Intermediation in a Multiproduct Banking Firm: An Econometric Study of U.S. Banks, 1979-82." C. Lawrence and R. Shay. eds., *Technological Innovation, Regulation, and the Monetary Economy.* Cambridge. Mass.. Ballinger, 1986.

Mester, Loretta J. "Efficient Production of Financial Services: Scale and Scope." *Business Review,* Federal Reserve Bank of Philadelphia, January/February 1987a, 15-25.

Mester, Loretta J. "A Multiproduct Cost Study of Savings and Loans." *Journal of Finance,* 42. June 1987, 423-445.

Murray, John D. and Robert W. White. "Economies of Scale and Economies of Scope in Multiproduct Financial Institutions: A Study of British Columbia Credit Unions." *Journal of Finance,* 38, June 1983. 887-901.

Scherer, F.M. *Industrial Market Performance and Economic Performance.* Chicago, Rand McNally, 2nd edition, 1980.

Demographic Influences on Household Growth and Housing Activity

By Glenn H. Miller, Jr.

The pace of home building in the United States during the 1980s differs sharply from that of the previous three decades. From 1950 to 1980, residential construction activity was a booming sector of the U.S. economy. In the 1980s, however, housing activity slowed considerably and its importance in the economy diminished. In light of this recent experience, an important economic question emerges: Will the slower growth of the 1980s continue in the years ahead, will it stabilize, or will the growth in housing regain the strength that it enjoyed during the earlier postwar years?

Many economic factors affect the growth of housing. Over the longer run, one of the most important of these factors is demographics—influences such as population growth, changes in the age structure, and changes in the rates of household incidence. These demographic influences—through their impact on household growth—played a key role in the strength of the housing industry from 1950 to 1980 and in the slowdown of the 1980s.

This article examines the past and future impact of demographic factors on the growth in the number of households and on housing activity in the United States. The article concludes that a projected further slowing of household growth through the end of the century is likely to be accompanied by further reduced growth in residential construction activity.

The article's first section discusses the determinants of household growth and its role in residential construction activity. The section also describes a framework for understanding how demographic factors combine to affect household growth. The following two sections use the framework to explain the accelerating growth in the number of households from 1950 to 1980 and the slower growth in the 1980s. The final two sections discuss the projection of increases in the number of households and their expected effect on future residential construction activity.

Glenn H. Miller, Jr., is vice president and economic adviser at the Federal Reserve Bank of Kansas City. Michael Grace, an assistant economist at the bank, assisted in preparation of the article.

TABLE 1
Number of households and residential construction activity, 1950-85
(Annual averages)

	Increase in the number of households	Residential investment	Private housing starts	Private housing starts plus mobile homes
	(thousands)	(bil. of 1982 $)	(thousands of units)	
1950-55	864	78.0	1,485	—
1955-60	985	87.6	1,317	—
1960-65	927	107.8	1,476	1,629
1965-70	1,193	108.9	1,373	1,691
1970-75	1,544	143.3	1,790	2,227
1975-80	1,931	158.9	1,716	1,976
1980-85	1,203	145.3	1,468	1,739

Sources: U.S. Bureau of the Census and U.S. Department of Commerce

Determinants of growth in the number of households

In the longer run, there is a demographic foundation for residential construction activity.[1] The pace of housing construction reflects growth in the number of households, because most new housing units are built to accommodate additional households, either directly or indirectly.[2] The growth in the number of households and the pace of residential construction activity shared a common pattern from 1950 to 1985, reaching their peaks in the 1970s, then declining in the 1980s (Table 1). Average annual increases in the total number of households reached a peak in the last half of the 1970s, as did annual averages for real residential investment (Chart 1).[3]

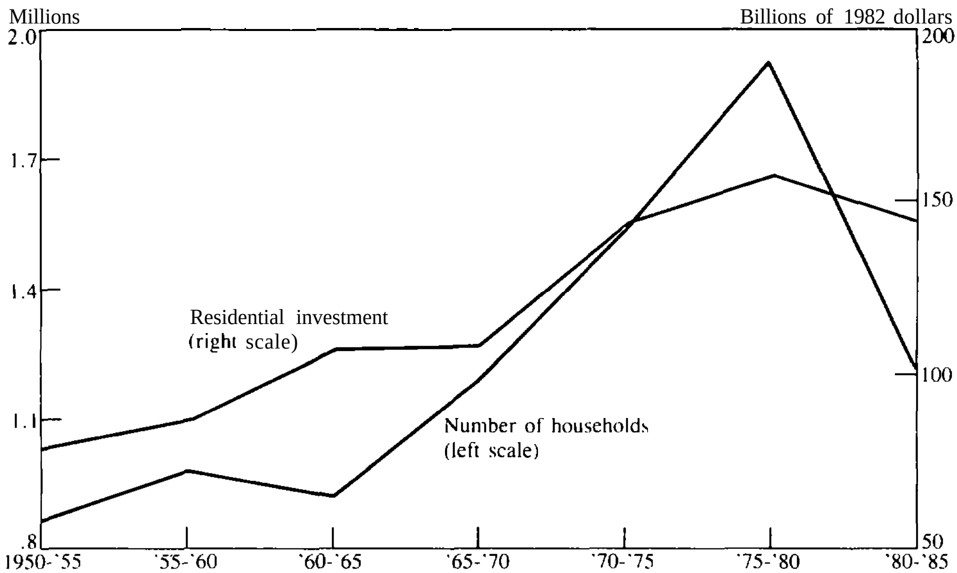
[1] Residential construction is subject to both short-run and longer run influences. Short-run movements in housing activity are usually caused by changes in financial variables, especially the availability and cost of home mortgage credit. Other financial factors include the cost and availability of funds to thrift institutions and of construction loans to home builders. Other economic factors can also influence short-run housing sector activity. These factors include construction costs, house prices, and the incomes and net worths of the households that occupy houses. Tax policy can also play an important role in the short-run effects on construction activity.

[2] Stuart A. Gabriel, "Housing and Mortgage Markets: The Post-1982 Expansion," *Federal Reserve Bulletin*, December 1987, p. 897.

[3] Other measures of housing activity behaved similarly. Total private housing starts, as well as starts plus mobile home shipments, peaked in the early 1970s and remained strong in the last half of the decade. As Table 1 shows, the increase in the number of households in a given period generally differs from the number of units built. According to the U.S. Bureau of the Census, "The number of housing units constructed [differs] from the increase in the number of households because of changes in the number of vacant units, the demolition of existing units, and conversions or mergers of units in existing structures." U.S. Bureau of the Census, Current Population Reports, Series P-25, No. 986, *Projections of the Number of Households and Families: 1986 to 2000*, Washington, D.C., 1986, p. 1. Housing activity not directly arising from growth in the number of households would include, for example, construction of second homes and of units built to accommodate internal migration.

CHART 1

Residential investment and increase in the number of households, 1950-85
(annual **averages)**



Source: Table 1

Demographic factors affect housing sector activity primarily through growth in the number of households. A household is defined as a person or group of persons occupying a housing unit, one of whom is identified as the householder.

Household growth is determined by population growth, the age structure of the population, and rates of household incidence. The age structure of the population is the distribution of the total population among various age groups. The rate of household incidence for any age group is the proportion of the population in that age group who are heads of households. For example, if there are 400 households for every 1,000 persons in a given age group, the household incidence rate for that age group is .400.

Household incidence rates are not just mechanical ratios, but result from the decisions and actions of persons. Household formation decisions depend on personal preferences and circumstances, and are often related to other decisions involving things like labor force participation and marital status.[4]

Households come into existence through the act of household formation: One or more persons establish separate living quarters by occupying a housing unit, which may be a house, an apartment, or a single room. The decisions and actions leading to household formation generally follow a life cycle. Children become young adults, leave their parents' homes, and set up their own households, thus requiring additional housing units.

[4] Louise B. Russell, ***The Baby Boom Generation and the Economy,*** Washington, ***D.C.,*** The Brookings Institution, 1982, p. 109.

They may do so as single persons who establish nonfamily households, or upon marriage when they establish family households.[5] After their children leave, parents remain a family household. Or because of death or divorce, one or both may return to living alone as nonfamily households.

If the patterns of decisions and actions remain about the same from generation to generation, then household incidence rates do not change much, and growth in the number of households is dominated by population growth and its age structure. But changes in decisions and actions relating to household formation can make changing household incidence rates a significant contributor to household growth.

A simple framework helps show how population growth, changes in the age structure, and changes in rates of household incidence contribute to growth in the number of households. The total number of households, H, is composed of the number of households in various age groups. If there are two age groups, H, and H_2, then $H = H_1 + H_2$. The number of households in a particular age group, H_1, depends on the population in that age group, P_1, and the rate of household incidence for that age group, h,. Thus, the total number of households at any given time can be written

$$H = P_1 h_1 + P_2 h_2.$$

The formula clearly shows that an increase or decrease in the number of households will occur if there is an increase or decrease in population, or if there is an increase or decrease in the household incidence rates. The number of households can also change if there is a change in the age structure of the population. To see this, suppose that the rates of household incidence are different for the two age groups $(h_2 > h_1)$ and that neither h_2 nor h, changes from period one to period two. Suppose also that the total population remains unchanged from the first to the second period. Finally, suppose that the age structure of the population changes such that the P_2 share of the total population increases, while the P, share declines. Because h_2 exceeds h,, this change in the age structure will cause the number of households to increase.

Changes in total population, population age structure, and household incidence rates may be reinforcing or offsetting. An illustration of combined effects is one in which total population increases, the age structure changes in such a way that growth is greater in age groups with higher incidence rates, and household incidence rates generally rise. These changes are reinforcing in the sense that all contribute toward increasing the growth in the number of households.

Growth in the number of households, 1950 to 1980

The rate of growth in the number of households increased from just under 2 percent per year in the 1950s to about 2.6 percent per year from 1975 to 1980 (Table 2). Throughout the period, the rate of growth in the number of households exceeded the rate of population growth. The average rate of total population growth declined from about 1.8 percent per year in the 1950s to about 1.1 percent per year in the period from 1975 to 1980 (Table 3). The divergence between population growth and growth in the number of households was especially large after 1965 (Chart 2).

5 A family household is a household maintained by a family of two or more persons related by birth, marriage, or adoption, and any unrelated persons. Nonfamily households consist of a person living alone or householders living with persons to whom they are not related. About nine-tenths of all nonfamily households are one-person households.

TABLE 2

Average annual increase in the number of households by type, 1950-87

	Percent			Number of households (in thousands)		
	Total	Family	Nonfamily	Total	Family	Nonfamily
1950-55	1.91	1.45	5.50	864	579	285
1955-60	1.98	1.48	5.19	985	635	350
1960-65	1.70	1.27	4.04	927	587	340
1965-70	2.00	1.47	4.49	1,193	724	469
1970-75	2.33	1.55	5.44	1,544	821	723
1975-80	2.59	1.40	6.42	1,931	797	1,134
1980-87	1.47	1.14	2.37	1,243	706	537

Source: U.S. Bureau of the Census

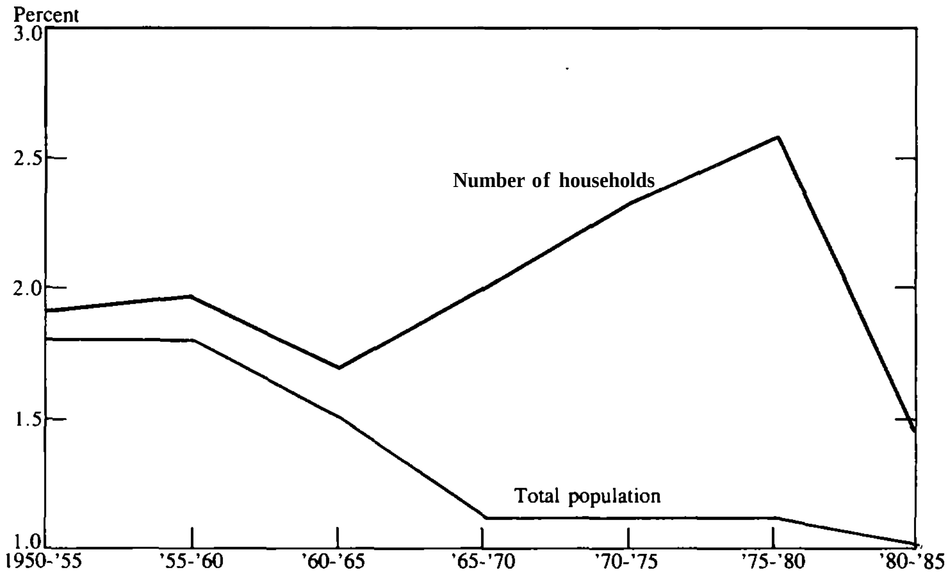
TABLE 3

Estimated and projected average annual percent change in population, by age, 1950-2000

	Age (years)					
	18-24	25-34	35-44	45-64	65 and over	Total
Estimated						
1950-55	−1.4	0.2	1.2	1.7	3.4	1.8
1955-60	1.6	−1.1	1.1	1.6	3.0	1.8
1960-65	5.2	−0.4	0.2	1.5	2.1	1.5
1965-70	4.4	2.5	−1.1	1.6	1.8	1.1
1970-75	2.7	4.9	−0.3	0.9	2.6	1.1
1975-80	1.7	3.9	2.7	0.3	2.7	1.1
1980-85	−1.1	2.5	4.6	0.2	2.2	1.0
Projected						
1985-90	−2.1	0.8	3.7	0.7	2.2	0.9
1990-95	−1.6	−1.4	2.2	2.5	1.4	0.8
1995-2000	0.7	−2.0	0.8	3.3	0.6	0.7

Source: U.S. Bureau of the Census

CHART 2
Total population and number of households, 1950-87
(average annual percent increases)



Source: Tables 2 and 3

Age structure and growth in the number of households

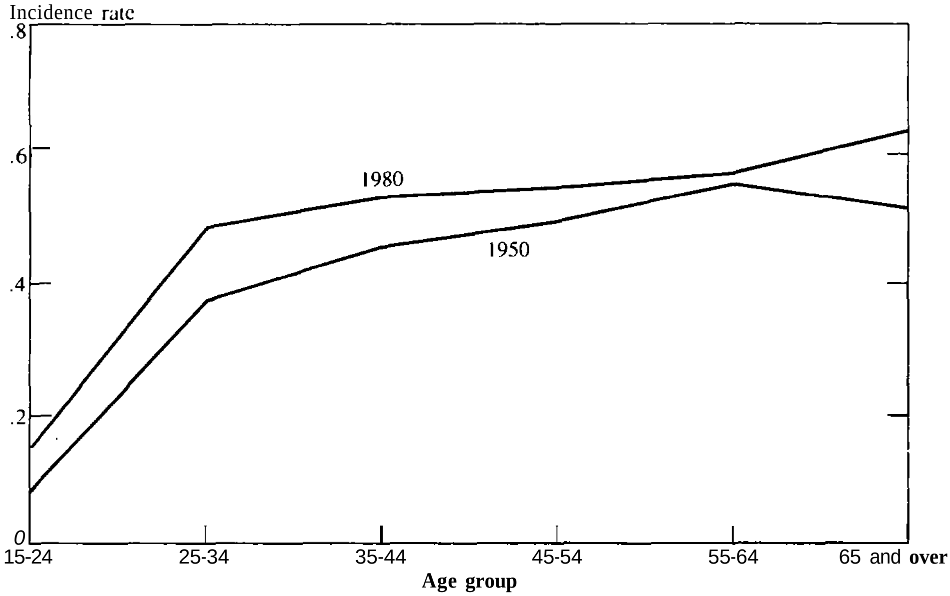
Faster growth in the number of households than in population stemmed largely from the changing age structure of the population, as changes in the age structure boosted household growth. The maturing of the baby-boom generation into age groups where household formation is more likely, and the relatively rapid growth of the population's oldest age group as life spans lengthened, each contributed to the age structure changes that increased growth in the number of households.

The coming of the baby-boom generation, those persons born in the years- 1946 through 1964, first increased the rate of population growth. Later, the baby boomers—a cohort both preceded and followed by smaller cohorts—became a major direct influence on the growth in the number of households.[6] As the baby boomers grew older, they moved into age groups where rates of household incidence are typically higher. The rate of household incidence for persons age 25-to-34 years is substantially higher than for those age 15-to-24, and the rate then rises more slowly through the rest of the age range (Chart 3).

Table 3 shows the inexorable progression of the baby-boom bulge through the population age structure. In the 1960s, the highest rates of population growth were in the 18-to-24 age group.

[6] The impact of the baby-boom generation on the U.S. economy and society has been, and is likely to continue to be, studied extensively. For a study of the importance of the baby-boom generation for the economy, see Louise B. Russell, *The Baby Boom Generation and the Economy.*

CHART 3
Household incidence rates by age group, 1950 and 1980



Source: Table 4

As the baby boomers matured, the 25-to-34 age group had the highest rate of growth in the 1970s. Population growth in each of these age groups first accelerated, then continued at much slower rates, or even declined, as the baby-boom generation moved through the age structure.

The rapid growth in the 1960s of the youngest age group initiated the significant contribution of the baby boomers to household growth. But the contribution was dampened because that age group's household incidence rate is the lowest in the adult population. As the large cohort of baby boomers moved from the youngest age group to the next older group, with its much higher incidence rate, the number of households increased sharply. Thus, the increase in the number of households with heads 25-to-34 years old due to their population growth alone was large in the last half of the 1960s and very large in the 1970s.

A second powerful factor joined the maturing of the baby boomers in changing the age structure —the fall in the death rate and the resulting longer life spans for older people.[7] Apart from the large age-specific percent increases produced as the baby boomers moved through the age structure, the largest average annual percent increases from 1950 to 1980 were for people age 65-and-over (Table 3). Because this age group had the highest rates of household incidence of any age group, the large increases in the population age 65-and-older contributed significantly to total growth in the number of households.

[7] Russell, p. 9.

TABLE 4
Rates of household incidence, by age group, 1950-87

Age	1950	1955	1960	1965	1970	1975	1980	1987
15-24	.095	.099	.104	.111	.119	.143	.154	.135
25-34	.373	.384	.424	.442	.462	.475	.492	.471
35-44	.453	.480	.480	.491	.510	.520	.540	.544
45-54	.500	.507	.529	.528	.524	.544	.556	.567
55-64	.559	.542	.550	.562	.579	.564	.576	.584
65 & over	.521	.544	.565	.594	.622	.628	.644	.637

Source: Calculations based on U.S. Bureau of the Census data

Incidence rates and growth in the number of households

The effects of changing age structure do not account for all of the increase in the number of households in the postwar period. Rising rates of household incidence also made a significant contribution, as changes occurred in decisions and actions affecting household formation. In the 1950s, marriage and childbearing occurred at earlier ages than before the war. But after the 1950s, trends toward delayed marriages and childbearing, more divorces, and more one-person households especially in the youngest and oldest age groups, all tended to increase household incidence rates.[8]

With few exceptions, rates of household incidence increased steadily from 1950 to 1980 for all age groups in the adult population (Table 4). Increases in household incidence were greatest in the youngest and the oldest age groups, thereby reinforcing the influence of changes in the population age structure.[9] Rising incidence rates brought faster growth in the number of households in every age group than can be explained by growth in population by age group.

To determine how much each of the demographic influences contributes to total household growth, Burnham Campbell developed a method to separate these influences. Campbell's method calculates the contribution to household growth made during any period by population growth and changes in the age structure, on the one hand, and changes in incidence rates, on the other hand.[10] In his analysis, the increase in the number of households in a given age group and over a given period that is solely attributable to the change in the size of the age group is called

8 Russell, p. 111.

9 Russell, pp. 92-93, 168.

10 Burnham O. Campbell, *Popularion Change and Building Cycles.* Urbana, Ill., Bureau of Economic and Business Research, 1966. especially Chapter 3. The methodology is also used in Russell, especially pp. 102-110.

"required additions." Required additions are calculated assuming that the incidence rate remains unchanged during the period. Thus required additions reflect the change in the number of households that would have occurred if the age group's population had changed as it did, but without a change in its incidence rate. To obtain the contribution made by a change in the incidence rate, required additions are subtracted from the actual change that occurred in the number of households in the age group, referred to as "actual additions." Summing over all age groups gives the actual additions, required additions, and additions due to changes in incidence for the total population."

In each period from 1950 to 1980, the total growth in the number of households, or actual additions, was greater than the increase in the number of households attributable to population growth and age structure alone, or required additions (Table 5). These estimates show the consistently important contribution of rising rates of household incidence. In every period, increased rates of household incidence were responsible for one-fourth or more of the total increase in the number of households. At the same time, the sharp increases in the number of required additions after 1965 show the substantial impact on the total growth in the number of households due to the changing age structure of the population.

The importance of the oldest and youngest age groups for growth in the number of households is clearly evident from Table 6. This table also shows the separate contributions of changing household incidence and of population growth by age group.[12]

[11] Russell, p. 105.

[12] Tables 5 and 6 differ from Tables 5-4 and 5-5 in Russell because of the use of data not available at the time of her study.

TABLE 5

Average annual additions to the number of households, 1950-87 (in thousands)

	Actual additions	Required additions	Additions due to changing household incidence
1950-55	864	625	239
1955-60	985	593	392
1960-65	927	601	326
1965-70	1,194	763	431
1970-75	1,543	1,037	506
1975-80	1,931	1,418	513
1980-87	1,243	1,420	−177

Source: Calculations based on U.S. Bureau of the Census data

Persons age 65-and-older contributed substantially to household growth from 1950 to 1980. The significant increase in the population age 65-and-older resulted in large increases in required additions for this age group, which had the highest of all incidence rates in any given year after 1950. At the same time, the increases in this group's incidence rate throughout the period reinforced its contribution to the total increase in the number of households.

The baby boomers' contribution to household growth is evident, too, as actual additions in the age group 25-to-34 years were extremely large from 1965 to 1980. The large numbers of required additions show the impact of the age structure change due to the maturing of the baby boomers into the 25-to-34 age group. Again, as was true for the oldest age group, the increases in required additions were reinforced by increases due to further rises in the household incidence rate for persons age 25-to-34 years.

Much of the rising rate of household incidence

TABLE 6

Additions to the number of households, by age group, 1950-80
(in thousands)

Age	1950-55	1955-60	1960-65	1965-70	1970-75	1975-80	Type of addition
15-24	36	410	854	946	1,475	735	Actual
	−56	288	644	640	509	276	Required
	92	122	210	306	966	459	Incidence
25-34	374	392	203	1,767	3,253	3,557	Actual
	92	−524	−192	1,263	2,840	2,924	Required
	282	916	395	504	413	633	Incidence
35-44	1,195	609	395	−199	51	2,119	Actual
	578	628	108	−637	−163	1,579	Required
	617	−19	287	438	214	540	Incidence
45-54	844	1,307	645	693	700	−262	Actual
	716	858	667	780	230	−545	Required
	128	449	−22	−87	470	283	Incidence
55-64	427	680	1,001	1,224	477	1,224	Actual
	685	544	799	902	789	968	Required
	−258	136	202	322	−312	256	Incidence
65 & over	1,442	1,527	1,538	1,537	1,760	2,284	Actual
	1,109	1,170	1,003	984	1,610	1,889	Required
	333	357	535	553	150	395	Incidence
Total	4,320	4,925	4,637	5,965	7,719	9,656	Actual
	3,124	2,964	3,029	3,932	5,815	7,091	Required
	1,196	1,961	1,608	2,033	1,904	2,565	Incidence

Source: Calculations based on U.S. Bureau of the Census data

in nearly all age groups from 1950 to 1980 was due to decisions and actions leading to rapid growth in the number of nonfarnily households, most of which are one-person households. The annual rate of growth in the number of nonfarnily households was much greater than the rate of growth for family households in each five-year period (Table 2). Over the three decades, the rate of family household growth averaged about one and a half percent per year, while nonfamily household growth averaged just over 5 percent per year. The rate of growth in the number of persons living alone was even higher—about 6 percent per year over the whole period.

The wide variation between the growth rates for family and nonfamily households substantially changed the composition of the total number of households. The total number of households in the United States nearly doubled from 43.6 million in 1950 to 80.8 million in 1980. In 1950, 89 percent of all households were identified as family households. By 1980, only 74 percent of all households were family households. Over the period, nonfamily households increased from 11 percent to 26 percent of the total, and persons living alone increased from 9 percent to 23 percent of all households. These increases reflect the trends toward later marriage, more divorces, and more older people living alone, which accounted for rising rates of household incidence.

Growth in the number of households in the 1980s

Growth in the number of households in the 1980s has differed sharply from that earlier in the postwar period, as annual percent increases in total households were considerably smaller in the 1980s (Table 2). The slowing was apparent for both family and nonfamily households, with a greater reduction in the rate of nonfamily household growth.

The slower household growth was due both to changes in the age structure of the population and to changes in rates of household incidence. While total population growth slowed only slightly in the 1980s, growth in the number of persons in the youngest age groups slowed considerably as the baby-boom generation continued to mature and was followed through the age structure by a smaller cohort (Table 3). Further, growth in the number of persons age 65-and-over in the 1980s has slowed somewhat compared with the 1970s. Rates of household incidence declined from 1980 to 1987 in the age groups 15-to-34 years and 65 years and over. Household incidence rates were higher in 1987 than in 1980 only for persons of ages 35 through 64, and only slightly higher for them (Table 4).

Estimates of how much of the slowing in household growth in the 1980s is due to changes in the age structure of the population and how much is due to changes in rates of household incidence are shown in Table 7. In contrast to the 1950-80 period, changes in household incidence rates in the 1980s acted to reduce the overall growth in the number of households rather than to add to it. If the population changes of the 1980s had been associated with the 1980 rates of household incidence, the number of households would have grown by about 9.9 million. But falling rates of household incidence made the actual increase in the number of households about 1.2 million less than that expected from population change alone.

The 1980-to-1987 experience was vastly different from earlier postwar years. For the six five-year periods from 1950 to 1980, required additions to households were never larger than actual additions in total, and almost never larger for any single age group. These patterns were clearly and sharply reversed in the 1980-to-1987 period. Where the youngest and oldest age groups had earlier been major contributors to total growth in the number of households, in the 1980s they were not. Falling rates of household incidence for those age 65-and-over and those age 34-and-under brought actual additions below required additions in those age groups. In the youngest age group, a fall in the rate of household incidence combined with a decline in population to produce a 21 percent drop in the number of households for that group. These changes more than offset the effect of slightly rising incidence rates in other age groups and brought total actual additions below total required additions. In the 1980s, age structure changes were no longer being reinforced by rising incidence rates, but were being offset by falling ones due to changing preferences, decisions, and actions relating to household formation.

TABLE 7
Additions to the number of households, by age group, 1980-87
(in thousands)

Age	Actual additions	Required additions	Additions due to changing household incidence	Addendum: Population change
15-24	−1,372	−656	−716	−4,262
25-34	1,998	2,898	−900	5,891
35-44	4,724	4,596	128	8,512
45-54	557	295	262	530
55-64	343	149	194	258
65 and over	2,454	2,660	−206	4,131
Total	8,704	9,942	−1,238	15,060
Annual Average	1,243	1,420	−117	2,151

Source: Calculations based on U.S. Bureau of the Census data

Projected growth in the number of households

The U.S. Bureau of the Census projects total population growth for the 1985-to-2000 period to be less rapid than it was earlier in the post-World War II period (Table 3).[13] As was true in earlier decades, however, growth in the number of households to the end of the century will depend on changes in age structure and rates of household incidence as well as on population growth.

Age structure changes

The age structure of the population in the years ahead will continue to be influenced strongly by the aging of the baby-boom generation and the maturing of the smaller cohort that follows it. As the baby-boom generation moved through the age structure in the postwar years, sharply increased rates of growth have been followed by slowing growth for the two youngest age groups (Table 3). Population declined in the 18-to-24 age group in the early 1980s, and projections show further decline until 1995 followed by only slow growth in the last half of the 1990s. The same pattern is projected for the 25-to-34 age group, but with a decade lag. Population in the two youngest age groups together is projected to decline by well over nine million persons, or about 13 percent, from 1985 to 2000.

Population growth in other age groups is not expected to pick up the slack. Slowing growth is projected for those persons age 35-to-44 years, with only very slow growth expected in the late 1990s. And the growth in the population age 65-and-over, whose contribution ranked close to that of the baby boomers in influencing earlier

[13] The U.S. Bureau of the Census projections of growth in the number of households discussed in this article are based on a "middle" population projections series prepared by the Bureau.

changes in the age structure of the population, is projected to slow steadily to the close of this century. With population in all age groups (except those age 45-to-64) declining or growing at much slower rates, this demographic support for growth in the number of households has weakened and will continue to weaken to the end of the century.

Incidence assumptions

Assumptions about future rates of household incidence are a key ingredient in projections of the number of households. Population growth by age group can be projected to the end of the century with some confidence, because those persons who will be 15 or older by the year 2000 have already been born, and mortality rates change slowly. Preferences, decisions, and actions affecting household formation shift more readily, however, with important effects on rates of household incidence and thus on the increase in the number of households.

In making the assumptions used in its projections of the number of households to the year 2000, the Census Bureau identified several demographic factors that have influenced past trends in household growth. The share of young adults maintaining their own households has increased and the share of young and middle-aged adults living in married-couple households has declined, because of increases in the proportion of persons who have never married or who were married and then divorced. In addition, changes in the age structure of the population tended to strengthen the effects of marriage and divorce on changes in household formation, as the baby boomers moved into the age groups where marriages were postponed and divorces more likely. Finally, the proportion of older persons maintaining their own households has increased. These factors, which are reflected in changing rates of household incidence, contributed to the earlier postwar acceleration in household growth.

More recently, however, some modifications of these trends have been occurring. The proportion of young adult men never married has been rising more slowly. There has been an increasing tendency for young adults to continue to live in their parents' homes. The divorce rate has fallen. And, the aging of the baby-boom generation brings its members into age groups where marriage is more widespread and divorce is less likely. Continuing moderation in rates of change in marriage, divorce, and living arrangements, due to the modifications just discussed, suggests a slower rate of growth in the number of households in the future.[14]

Projections of the number of households

The Bureau of the Census projects slower growth in the number of households to the end of the century than occurred in earlier decades.[15] Average annual increases shown in Table 8 exhibit the slower projected growth, compared with the earlier growth shown in Table 2.

Each of the Bureau of Census projections series takes account of age structure changes while reflecting different demographic assumptions affecting rates of household incidence. Projections Series I reflects "the demographic assumption that the era of rapid change in marriage and divorce may have come to an end, and conse-

[14] U.S. Bureau of the Census, Current Population Reports, Series P-25, No. 986, *Projections of the Number of Households and Families: 1986 to 2000,* Washington, D.C., 1986, p. 5.

[15] The projected increases in the number of households were designed to be plausible. yet only illustrative of long-run changes. The projections form smooth trends but actual future changes are unlikely to be smooth "because of short-term fluctuations due to various social and economic factors." The projections assume the absence of major catastrophes such as general war, and of large unexpected changes in underlying demographic trends. *Projections of the Number of Households and Families: 1986 to 2000,* p. 5.

TABLE 8
Projected average annual increase in the number of households, 1985-2000

	Percent		Number of households (in thousands)	
	Series I	Series II	Series I	Series II
1985-90	1.41	1.62	1,230	1,416
1990-95	1.05	1.29	977	1,216
1995-2000	0.87	1.12	852	1,125

Source: U.S. Bureau of the Census

quently, that householder proportions will remain constant" at their 1985 levels.[16] With the rates of household incidence in effect at the beginning of the projections period kept unchanged, Series I reflects the results of projected change in population age structure alone and is essentially equivalent to projections of required additions. Projections Series II assumes further changes in household incidence after 1985, due to continued moderation of underlying trends in marriage and divorce.[17]

Series I, which holds incidence rates constant at 1985 levels, projects required additions that are small compared with the actual additions of earlier decades. Annual percent increases for five-year periods are below any experienced in the 1950-to-1985 period, and the increase in the number of households projected for the last half of the 1990s is below that of any earlier five-year period shown in Table 2. Thus, if growth in the total number of households is not to slow drastically, rates of household incidence must make some contribution.

The Series II projections show the results when changing household incidence rates contribute to increases in the number of households. With its assumption of moderate changes in factors affecting household incidence, Series II shows faster growth in the number of households than does Series I. But even with this contribution, household growth to the century's end is still slower than for most of the postwar period. Moreover, the projected increase in the number of households in the 1990s is also very small in Series II, well below that of the 1960s and 1970s.

Growth in the number of households is thus projected to be substantially less to the year 2000 than that experienced earlier. Indeed, even slower future growth could result if changes in household incidence were to have a negative influence on the increase in the number of households, such as occurred from 1980 to 1987.

16 *Projections of the Number of Households and Families: 1986 to 2000*, p. 5.

17 The U.S. Bureau of the Census produced three different projections of household growth, called Series A, B, and C. Series B and C are discussed in this article as Series II and I, respectively. While Series C assumes unchanged incidence rates, Series A and B both assume further changes in incidence rates. Series B, which assumes greater moderation of earlier trends in incidence rates, produces projections of slower household growth than does Series A. At the time of their publication, the U.S. Bureau of the Census judged Series B to be the most plausible set of projections. Because of that judgment, and in light of the slowing of household growth in the 1980s, Series A is not discussed here.

Implications for residential construction activity

An important result of the projected slowing of growth in the number of households is slower growth to the end of the century in residential construction activity, compared with earlier postwar growth. While forecasts of the number of housing units produced or of the amount of residential investment cannot be made from household growth projections alone, the influence of other factors would have to be substantial to offset the effect of significantly slower growth in the number of households.

One way to look at the possible impact of slower household growth and hence of reduced residential construction activity is within a projection of total economic activity. A Bureau of Labor Statistics (BLS) projection of the U.S. economy in 2000 estimates real GNP growth from 1986 to 2000 at about the average rate of the previous 15 years.[18] This projection incorporates a substantial slowing in residential construction activity and suggests some shifting of gross private domestic investment from the residential to the nonresidential component.[19]

Within this BLS projection, real residential investment is projected to grow at 0.4 percent per year for the rest of the century, well under the 1.3 percent annual average growth from 1972 to 1986. Residential investment in 2000 would be 3.9 percent of total real GNP, compared with 5.3 percent in 1986 and 6.2 percent in 1972. The projected slowdown in residential construction growth is due to both cyclical and demographic factors. With regard to demographic factors, the BLS says that "the formation of new households is projected to slow dramatically during the 1990s, pulling down the level of housing starts over the projection period."[20] Moreover, the BLS mid-range projections presented here are based on assumed growth in the number of households faster than that in either of the projections series discussed earlier. Unless offset from other sources, additional weakness in its demographic foundation might further reduce housing activity.

Summary

While residential construction is affected by a number of factors, the relationship between growth in the number of households and residential construction activity is an important one. Growth in the number of households is the fundamental support underlying growth in housing construction over the longer run. Household growth has slowed in the 1980s from earlier post-World War II decades, and projections to the year 2000 by the Bureau of the Census show further slowing. The slower growth in the number of households will almost surely be reflected in residential construction activity. Thus, projected slower growth to the end of the century in the number of households suggests slower growth in residential construction, leading possibly to slower total economic growth or to a change in the composition of output.

---

18 Norman C. Saunders, "Economic Projections to the Year 2000," *Monthly Labor Review*, September 1987, pp. 10-18. This article's discussion is based on the BLS mid-range projection.

19 Such a shifting might be welcomed by some observers, as a number of analysts have argued that too many resources have gone into residential investment at the expense of other fixed capital investment in the United States. For example, see Edwin S. Mills, "Has the United States Overinvested in Housing?" *Journal of American Real Estate and Urban Economics Association*, Vol. 15, No. 1, Spring 1987, pp. 601-616.

20 Saunders. pp. 14-15

Economic Review
Federal Reserve Bank of Kansas City
Kansas City, Missouri 63198
September/October 1988, Vol. 73, No. 8