Federal Reserve Bank of New York

Economic Policy Review

June 1998
Volume 4 Number 2

The ECONOMIC POLICY REVIEW *is published by the Research and Market Analysis Group of the Federal Reserve Bank of New York. The views expressed in the articles are those of the individual authors and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System.*

FEDERAL RESERVE BANK OF NEW YORK

ECONOMIC POLICY REVIEW ADVISORY BOARD

Federal Reserve Bank of New York Economic Policy Review

Table of Contents

June 1998
Volume 4 Number 2

Despite investors' willingness to hold a variety of financial assets and risks, a significant share of interest rate options exposures remains in the hands of dealers. This concentration of risk makes the interest rate options market an ideal place to explore the effects of dealers' dynamic hedging on underlying markets. Using data from a global survey of derivatives dealers and other sources, this article estimates the potential impact of dynamic hedging by interest rate options dealers on the fixed-income market. The author finds that for short-term maturities, turnover volume in the most liquid hedging instruments is more than large enough to absorb dealers' dynamic hedges. For medium-term maturities, however, an unusually large interest rate shock could lead to hedging difficulties.

This article is the first formal investigation of consumer attitudes that compares the forecasting power of the University of Michigan's Index of Consumer Sentiment and the Conference Board's Consumer Confidence Index. The authors find that measures available from the Conference Board have both economically and statistically significant explanatory power for several categories of consumer spending. By contrast, measures available from the University of Michigan generally exhibit weaker forecasting power for most categories of spending. As part of their analysis, the authors examine the ways in which the surveys underlying these measures differ and test whether certain types of survey questions are particularly important for predicting consumer spending.

A major feature of globalization has been the enormous increase in international flows of goods and services: countries are now trading much more with each other. In this article, the authors demonstrate the greater role vertical specialization is playing in these increased flows. Vertical specialization occurs when a country uses imported intermediate parts to create a good it later exports—that is, the country links sequentially with other countries to produce a final good. Deriving evidence from four case studies as well as OECD input-output tables, the authors reveal that vertical specialization has accounted for a large and increasing share of international trade over the last several decades. They also note that because the trends encouraging vertical specialization—lower trade barriers and improvements in transportation and communications technologies—are likely to continue, this type of international trade should become even more prevalent in the next century.

Policy Rules and Targets: Framing the Central Banker's Problem

Stephen G. Cecchetti

Central bank policymakers are not primarily random number generators.[1] Reading both the financial press and the work of academics, however, one might get the opposite impression. Reporters (and the readers of their stories) seem to attach considerable importance to each Federal Open Market Committee policy decision. Academic work on the impact of central bank policy gives a similar impression, as statistical procedures produce a time series of pure white noise innovations that are labeled "policy shocks."[2] But central bankers expend substantial energy attempting to tailor their actions to current economic conditions. In other words, policymakers are reacting to the environment, not injecting noise.

But what is central bank policy anyway? The policymaker's problem can be characterized in the following way. Using an instrument such as an interest rate, together with knowledge of the evolution of the economy (aggregate output and the price level), the policymaker seeks to stabilize output and prices about some path that is thought to be optimal. In carrying out this goal, the policymaker must often trade off variability in output for variability in prices because it is generally not possible to stabilize both. This process yields what most people would call a policy rule, that is, a systematic rule for adjusting the quantity that the central bank controls as the state of the economy fluctuates. In other words, the study of policy should focus on the systematic portion of policymakers' actions, not the shocks.

In this essay, I discuss a number of conceptual and practical issues associated with viewing policymaking in this analytical framework. These issues include the implications for policymaking of the slope of the output-inflation variability trade-off, the influence of various types of uncertainty on the policymaker's problem, the consequences of the fact that the nominal interest rate cannot fall below zero, and possible justifications for interest rate smoothing.

Stephen G. Cecchetti is executive vice president and director of research at the Federal Reserve Bank of New York.

Although my intention is to raise, rather than resolve, key questions concerning the formulation of a policy rule, I do offer important new evidence on one point. This concerns the potential consequences of the move by many central banks to adopt some form of price-level or inflation targeting. In taking this approach, central banks are implicitly altering the relative importance of inflation and output variability in their objectives, increasing the weight they attach to the former relative to the latter. But the data suggest that the output-inflation variability tradeoff is extremely steep, implying that an effort to decrease inflation variability modestly could lead to a significant increase in output variability. Thus, policymakers considering pure inflation targeting should be aware that their change in emphasis could have undesirable side effects.

An Analytical Framework for Policy Formulation

As I suggested in the introduction, central bank policy can be thought of as the solution to a problem in which the policymaker uses an interest rate to stabilize the variability of output and prices about some path. A truly complete

Central bank policy can be thought of as the solution to a problem in which the policymaker uses an interest rate to stabilize the variability of output and prices about some path.

description of the policymaker's problem begins with an intertemporal general equilibrium model based on a social welfare function (tastes), production functions (technology), and market imperfections that cause nominal shocks to have real effects (nominal rigidities). The goal would be welfare maximization.

I do not propose to delineate the fully specified problem. Instead, I begin with a commonly used quadratic loss function that might be a second-order approximation to the objective function in this more detailed problem.[3] The policymaker seeks to minimize the discounted sum of squared deviations of output and prices from their target paths. The general form of such a loss function can be written as

$$L = E_t\left(\sum_{i=0}^{h} \beta^i \{\alpha[p_{t+i} - p^*_{t+i}]^2 + (1-\alpha)[y_{t+i} - y^*_{t+i}]^2\}\right), \quad (1)$$

where p_t is the (log) aggregate price level, y_t is the (log) aggregate output, p^* and y^* are the desired levels for p and y, β is the discount factor, h is the horizon, α is the relative weight given to squared price and output deviations from their desired paths, and E_t is the expectation conditional on information at time t.[4] The loss function provides the policymaker with information about preferences over different paths for the variance of output and prices.

A complete formulation of L requires description of p^* and y^*. I will focus on the desired price path, ignoring issues concerning y^*.[5] Here we encounter the following question: Should the objective be a price-level path or an inflation rate? The first of these, *level targeting*, would dictate that

$$p^*_t = p^*_{t-1} + \pi^* = \pi^* t, \quad (2)$$

where π^* is the desired steady level of inflation. That is, the optimal price level this period is the optimal level last period plus some optimal change (which may be zero). The alternative, *rate targeting*, is

$$p^*_t = p_{t-1} + \pi^*, \quad (3)$$

where the current target price level is just the last period's *realized* price level plus the optimal change.

The difference between price-level and inflation rate targeting is the path for the variance of prices. Level targeting implies more volatile short-horizon prices and less volatile long-horizon prices than does rate targeting. To see this, simply note that equation 3 implies that

$$p^*_t = \pi^* t + \sum_{i=0}^{\infty} (p_{t-i} - p^*_{t-i}),$$

which can be a random walk.[6]

The description of the loss function is now complete. It is a function of the parameter vector $\theta = \{\alpha, \beta, h, \pi^*\}$. The values of each of these will depend on the underlying economic structure, that is, tastes and technology. The preference for paths with greater or lesser degrees of variability in output relative to variability in prices, as embodied in the loss function, depends on the fundamental reason that these things are costly. The same is true of the desired steady level of inflation, π^*.

The policymaker's problem cannot be solved without knowledge of the dynamics of output and prices as functions of the policy control variable and the stochastic forcing process driving the economy. These relations, which are taken as constraints in the optimization problem, describe the structure of the economy. For the purposes of the current discussion, I will assume that the central bank policy is carried out using an interest rate, r_t,[7] and that the innovations to the economy come from a series of real and nominal shocks (that is, aggregate demand and aggregate supply shocks), which can be written as ε_t.[8] The reduced form for the evolution of output and prices can then be written as

(4)
$$\begin{bmatrix} y_t \\ p_t \end{bmatrix} = A(L) \begin{bmatrix} \varepsilon_t \\ r_t \end{bmatrix},$$

where $A(L)$ is an $(n+1) \times 2$ matrix of (possibly infinite-order) lag polynomials in the lag operator L.[9] The coefficients in $A(L)$ describe a reduced form of the economy. For the moment, I will ignore the fact that $A(L)$ is likely to change when the policy rule changes.[10]

We can now characterize the policymaker's problem as choosing a path for r_t that minimizes the loss (equation 1), with either equation 2 or equation 3 substituted in for p^*, subject to equation 4. The result is a policy rule, which I will write as

(5)
$$r_t = \phi(L)\varepsilon_t,$$

where $\phi(L)$ is a (possibly infinite-order) lag polynomial.[11] This path for interest rates as a function of the innovations to the economy (which could be written as differences in the observable quantities) is the policy rule. Significantly, $\phi(L)$ is a function of the parameters θ, as well as the coefficients in $A(L)$ and the covariance matrix of ε, Σ.

I would like to emphasize that the preferences for the evolution of output and price variability, as well as the optimal steady inflation rate π^*, are *inputs* into the policymaker's problem.[12] In practice, I expect that these inputs would be dictated by some legislative or executive body in the government, as they are in some countries (although not in the United States). Given this objective (the loss function) and a model for the evolution of output and prices (the economy), the policymaker chooses a rule that governs the path of the control (the interest rate).[13]

Conceptual Issues and Practical Considerations

The preceding section provides an analytical framework for understanding the policymaker's task, or "problem." In this section, I use this framework to explore several issues relating to policy formulation. Although I leave many questions unanswered, my approach casts new light on some old problems and suggests directions for future research.

I will consider five issues. I begin by exploring the nature of a target. I proceed to a discussion of the practical problems posed by the apparent steepness of the output-inflation variability trade-off and consider how it might influence decisions. This is followed by a general discussion of how uncertainty affects policymaking. Next, I discuss how the nonlinearity created by the fact that the nominal interest rate cannot fall below zero influences the policy rule. Finally, I explore the issue of interest rate smoothing.

Policy Targets

If we accept the view that policy formulation is essentially the solution to the analytical problem of choosing a path for a control variable given a loss function, then how should we interpret the current debate over the proper choice of a policy target, and the advisability of targeting in general? I will explore two ways of addressing the issue of targets. The first is purely technical, and the second has to do with the way policymakers might portray their objective to the public. Technically, the first-order conditions (or Euler equations) to the loss minimization problem I describe above may be interpreted as producing a type of targeting regime. To see this, consider the case examined in detail by Svensson (1996b). He considers pure inflation rate targeting and a loss that is independent of output variation ($\alpha = 1$). The first-order condition of this problem implies setting the path for expected inflation, $E_t\pi_{t+i}$, as close to the optimal value, π^*, as possible. Svensson refers to this as "inflation forecast targeting."[14]

The only case I can see for intermediate targeting is that it contributes to policy transparency.

This analysis can then be used to justify public statements by policymakers that they are targeting inflation forecasts, as a rhetorical device that substitutes for the more complex and less accessible statements that would be needed to describe their entire procedure.

Ball's (1997) analysis suggests another justification for targets. The argument is that the loss minimization procedure of the type described in the preceding section is too difficult to explain to the population at large (and possibly their elected representatives as well), and so will not lead to policy that is transparent enough to ensure the proper level of accountability.[15] But a pure inflation targeting rule is easy to explain and, more important, easy to understand and monitor. As a result, if the solution to the complex problem can be approximated by a simple rule, there may be substantial virtue in adopting the approximate solution.

A related issue concerns the usefulness of intermediate targets. Over the last half-century or so, many monetary economists have advocated targeting various monetary aggregates. Consider the example of M2.[16] Researchers do not claim to care about M2 for its own sake, nor do they claim that central banks can control it exactly. Therefore, M2 is neither a direct objective nor an instrument. Instead, it is somewhere in between—an *intermediate* target—and the target path would again be akin to the first-order conditions of the optimal control problem.

I find it difficult to make an argument for monetary aggregates as intermediate targets. To see why, consider the case in which the policymaker controls an interest rate and cares about the price level ($\delta = 1$). To control the objective, the policymaker must know how prices respond to changes in the exogenous environment (the response of p_t to ε_t) and how the objective responds to changes in the instrument. But how does an intermediate target such as M2 help? Clearly, if the relationship between interest rates and M2 and that between M2 and prices are both stable and precisely estimable, then looking at the two relationships separately yields no advantage. In some instances, estimating the impact of interest rates on M2 and the impact of M2 on prices separately might give a more reliable estimate of the product of the two, but such instances would surely be rare. If M2 helps forecast prices, then it will be included in the model. But there is substantial evidence, some of which is in Cecchetti (1995), that reduced-form inflation forecasting relationships are very unstable even if they include M2 or any other potential intermediate target.[17]

As a result, the only case I can see for intermediate targeting is that it contributes to policy transparency. Svensson (1996b) describes an ideal intermediate target that "is highly correlated with the goal, easier to control than the goal, easier to observe by both the central bank and the public than the goal, and transparent so that central bank communication with the public and public understanding and public prediction of monetary policy are facilitated" (pp. 14-5). But since monetary aggregates cannot be closely controlled, are only weakly correlated with both output and inflation over horizons of months or even several years, and have changing definitions that make them difficult to explain, they fail to meet most of Svensson's criteria.

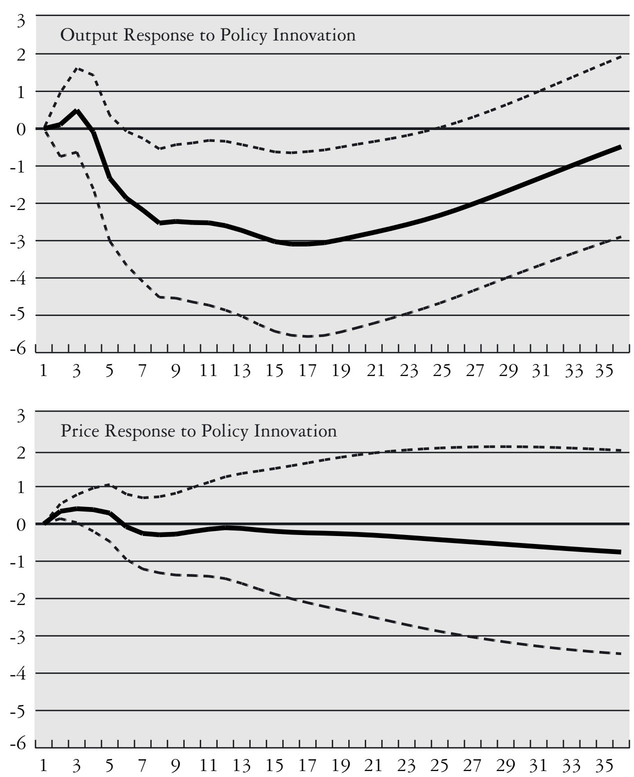
The Output-Inflation Variability Trade-Off

One of the most important practical issues facing policymakers concerns the output-inflation variability trade-off. To measure this trade-off, I turn to some earlier empirical estimates of the impact of central bank policy on output and prices (Cecchetti 1996). In effect, these estimates are the lag polynomials in equation 4 associated with r_t.[18] Chart 1 plots the *impulse response functions*, or dynamic reactions of prices and output to innovations (ε_t's), on the same vertical scale. The most important point to note is that the impact of policy innovations on output is both large and immediate. By contrast, policy affects prices only very slowly, and by much more modest amounts. Furthermore, the precision of the estimates is quite poor.

It is important to keep in mind that standard econometric methods, such as those employed here, assume that parameters are constant over significant historical periods. That is, the vector autoregression (VAR) method used to estimate the response of output and prices to interest rate movements presumes that these reactions are fixed over the 1984-95 sample used in the estimation. Numerous things can cause these relationships to change. As I emphasize in Cecchetti (1995), shifts in central bank policy regimes during the 1970s and 1980s will result in changes in the impulse response functions plotted in Chart 1. In the context of the current discussion, this means that I can reliably measure the output-inflation variability trade-off given the policy regime that was in place over the decade ending in 1995. I cannot reliably estimate the impact of dramatic changes in the policy regime on the trade-off.[19]

Chart 1

RESPONSE OF OUTPUT AND PRICES TO POLICY INNOVATIONS



Note: The dotted lines represent standard deviation bands of ±2.

To continue, with the aid of a very simple model, these estimates can be used to give some sense of the shape and slope of the output-inflation variability trade-off. Consider the simple one-period case in which the horizon in the policymaker's loss function (h) is zero, the discount factor (β) is irrelevant, target levels of output (y^*) and prices (p^*) are zero (in logs), and the structure of the economy is such that

(6) $$y_t = \gamma r_t + d_t - s_t,\ \gamma < 0 \quad \text{and}$$

(7) $$p_t = -r_t + d_t + s_t,$$

where d_t and s_t are aggregate demand and aggregate supply shocks. Demand shocks raise both output and prices, while supply shocks move them in opposite directions. I assume that the two types of shocks are uncorrelated and that the variance of the supply shocks is normalized to one, while the variance of the demand shocks is given by σ_d^2.[20] The parameter γ is a measure of the impact of policy innovations on output relative to their impact on prices. The example is meant to represent the medium-horizon impact of policy on the variables of interest. In this simple linear case, the policy rule will be

(8) $$r_t = ad_t + bs_t.$$

Equation 8 implies that

(9) $$\sigma_y^2 = (\gamma a + 1)^2 \sigma_d^2 + (\gamma b - 1)^2 \quad \text{and}$$

(10) $$\sigma_p^2 = (1 - a)^2 \sigma_d^2 + (1 - b)^2.$$

Minimizing the loss function

(11) $$L = \alpha \sigma_p^2 + (1 - \alpha)\sigma_y^2 \quad \text{yields}$$

(12) $$\alpha = \frac{\alpha - \gamma(1-\alpha)}{\alpha + \gamma^2(1-\alpha)} \quad \text{and}$$

(13) $$b = \frac{\alpha + \gamma(1-\alpha)}{\alpha + \gamma^2(1-\alpha)}.$$

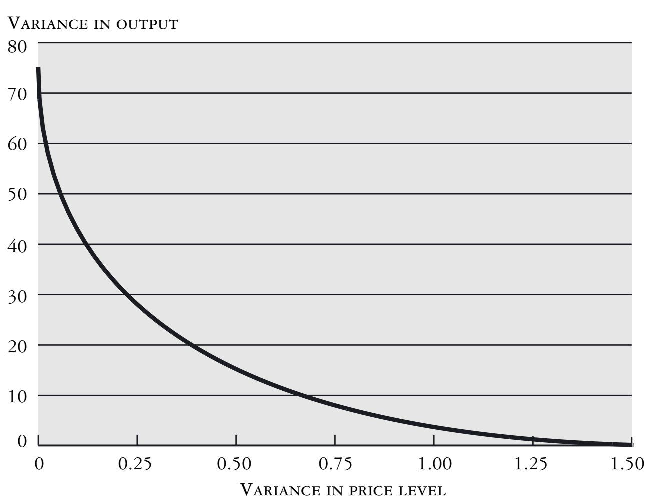
Substituting these into the variance expressions 9 and 10 yields σ_p^2 and σ_y^2 as functions of α, γ, and σ_d^2.

Using the monthly data from Cecchetti (1996), I can now compute an approximate output-inflation variability frontier. From Chart 1, I approximate γ as the mean of the response of output to the mean of the response of prices to an interest rate shock. The result, the average value over a thirty-six-month horizon for the 1984-95 period, is -6.74. The medium-run horizon chosen for these calculations is relevant to policymaking. In a full multiperiod framework, the definition of γ would be more complicated, but its interpretation would remain the same.

Once I determine σ_d^2, the ratio of the variance of demand shocks to the variance of supply shocks, then varying α allows construction of the frontier. Setting σ_d^2 to 0.46 forces the frontier to pass through the value implied by the data (the ratio of output to price variability is approximately 3.72), and normalizing the variance of the detrended log price level in the data to be equal to one gives Chart 2. The "X" marks the value implied by the data. Note that the 1984-95 data suggest that policymakers were operating as if α were approximately 0.93. This is consistent with the importance attached to low and steady inflation over this period.

Chart 2

THE INFLATION-OUTPUT VARIABILITY TRADE-OFF
1984-95



Source: Author's calculations.

Significantly, Chart 2 shows that the trade-off is extremely steep. Reducing inflation variability entirely by setting $\alpha = 1$ creates an extremely high level of variability in real output. In fact, moving from the historically observed point where the ratio of output to inflation variability is 3.72, setting σ_p^2 to zero would increase the variability of output by a factor of more than twenty! By contrast, reducing output variability from 3.72 to zero increases price variability from 1.0 to 1.65. This finding is not a consequence of the simplicity of the example, but rather of the fact that γ is so large. It is straightforward to show that the maximum value of σ_y^2, at $\alpha = 1$, is γ^2 times the maximum value of σ_p^2, at $\alpha = 0$, minus one. That is to say, the points where the line in Chart 2 intersects the x- and y-axes are solely determined by the size of the ratio of policy innovations' impact on output to policy innovations' impact on prices.[21]

Someone who cares about output variability is made substantially worse off by a decision to target the path of the price level.

This result has important implications for the current policy debate. As many central banks move toward some form of price-level or inflation targeting, they are implicitly changing the relative importance of output and inflation variability in their objective function, raising α toward one. From a purely pragmatic point of view, someone who cares about the aggregate price path loses little by allowing α to be less than one. The reverse, however, is emphatically not true. Someone who cares about output variability is made substantially worse off by a decision to target the path of the price level. As a result, when considering policies based on prices alone, policymakers must be very cautious and ask whether they really care so little about output and other real quantities.

Because the estimate of γ plays a crucial role in these conclusions, some comment on its statistical properties is in order. Unfortunately, the estimate is extremely imprecise, with an estimated standard error in excess of 18.[22] This difficulty almost surely stems from the relative stability of inflation during this period and the small and imprecisely estimated response of aggregate prices to policy innovations.[23] The immediate implication is that it is very difficult to be confident of the slope of the variability trade-off. It could be somewhat better, but it could also be substantially worse. A natural reaction to this is to examine the implications of uncertainty for the optimal policy rule.

Uncertainty

How does uncertainty affect policy? Of the numerous types of uncertainty that might influence central bank policy decisions, two forms are examined here: uncertainty about the impact of policy changes (on output and prices) given the model of the economy, and uncertainty about the model itself.[24]

It is straightforward to consider the first of these, which is the sampling error from the estimation of the reaction of prices and output to changes in the policy instrument. In the simple example here, this is just the variance of the estimated γ, which I will call $\sigma^2_{\hat{\gamma}}$. Brainard (1967) originally noted that this type of uncertainty leads to caution in that policy rules imply smaller reactions.[25] In this simple example, inclusion of $\sigma^2_{\hat{\gamma}}$ implies that the policy parameters *a* and *b* become

$$(14) \qquad a = \frac{\alpha - \hat{\gamma}(1-\alpha)}{\alpha + (\hat{\gamma}^2 + \sigma^2_{\hat{\gamma}})(1-\alpha)} \quad \text{and}$$

$$(15) \qquad b = \frac{\alpha + \hat{\gamma}(1-\alpha)}{\alpha + (\hat{\gamma}^2 + \sigma^2_{\hat{\gamma}})(1-\alpha)}.$$

Reactions to a given shock are now smaller.

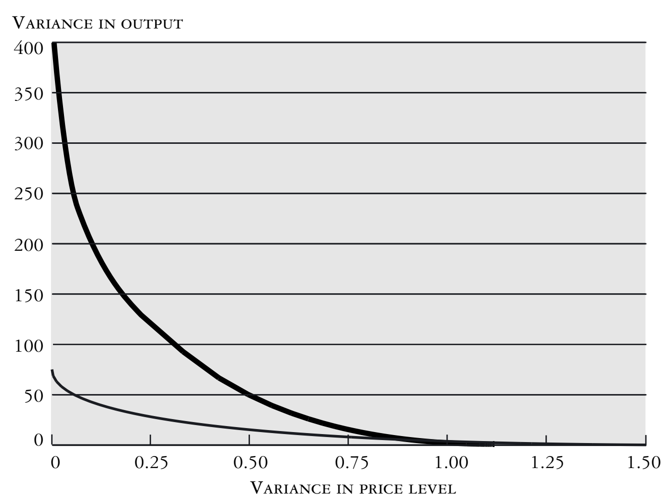
In a more realistic, multiperiod context, accounting for parameter uncertainty can be very difficult. Is it likely to be worth the trouble? To get some sense of the impact of parameter uncertainty, I use the results from the previous exercise. If, as I found there, the estimate of γ has a standard error equal to 18.7, then the variance will be 350! The results, plotted in Chart 3, suggest that the impact is huge: the variability frontier shifts out dramatically. Employing the same methods as in Chart 2, I have computed the implied value for σ^2_d so that the inflation-output variability frontier again goes through the point implied by the data. This occurs when σ^2_d equals 0.08 and α equals 0.89, compared with 0.46 and 0.93 in the certainty case. Interestingly, assuming that policymakers have minimized the simple loss function (equation 11) in the presence of uncertainty leads one to conclude that aggregate demand shocks have been substantially less important.

But the real implication of uncertainty is that the frontier is now substantially steeper, and the reaction function parameters *a* and *b* are significantly smaller. In fact, taking account of the changes in both the parameters and the average size of a typical demand or supply shock, I conclude that uncertainty leads to reactions that are on the order of one-twenty-fifth what they were in the certainty case.

What about model uncertainty? There are essentially two problems here. First, past history may not be a reliable guide to the impact of future policy actions, because underlying economic relationships, which policymakers had previously been able to exploit, may change. Such changes could be brought about by policy itself. It is this point, first noted by Lucas (1976), that has driven many macroeconomists to work on dynamic general equilibrium models with well-articulated microeconomic

Chart 3

Impact of Uncertainty on the Variability Trade-off



Source: Author's calculations.

foundations. But these efforts are still at too early a stage to be of practical use.

Second, there is little agreement over the true structural model of the economy. McCallum (1997) argues convincingly that, as a result of this lack of consensus, a policy rule should be robust to the possibility that numerous models are correct. In the context of the analytical framework presented in this essay, identifying such a rule would mean exploring the implications of various $A(L)$ s, each of which corresponds to a different model. The object would be to look for a rule that would perform well for a wide range of choices. One method for handling model uncertainty would be to treat it as variance in the estimate of the parameters in $A(L)$.[26] Overall, however, I am forced to conclude that we know very little about how to solve this problem.

THE ZERO NOMINAL INTEREST RATE FLOOR

What average inflation level should the policymaker target? There are two parts to this question. First, what is the optimal level of inflation, π^*? Second, should policy allow the average realized level of π to deviate from this level?

I argued above that π^* should be dictated to the central bank by social welfare considerations. Quite a bit of work has been done on the possible labor market benefits of modest inflation, suggesting that the optimal level may exceed zero. Most recently, Akerlof, Dickens, and Perry (1996) and Groshen and Schweitzer (1997) consider whether small positive levels of aggregate inflation can facilitate real adjustments in the presence of an aversion to nominal wage declines. But Feldstein (1996) contends that the tax distortions created by inflation reduce the level of output permanently, an argument that suggests π^* may even be negative.[27] Overall, we await further research for the definitive resolution of this issue.

There is one dominant argument for why policymakers might choose to allow average inflation to deviate systematically from the optimal level. The argument, raised in Summers (1991), concerns the case in which π^* is zero, and focuses on the fact that the nominal interest rate cannot fall below zero. In fact, any choice of π^* bounds the real interest rate. Summers goes on to note that in the historical record, the real interest rate (at least ex post) has often been negative. But if central bank policymakers successfully target zero inflation, then the fact that the nominal interest rate cannot be negative means that the real interest rate must always be positive. In essence, this restricts the ability of the policymaker to respond to certain shocks. The control problem as it is described above does not explicitly consider the fact that r_t is bounded at zero. As a result, there will be realizations of ε_t in which the policy rule (equation 5) would imply negative values for the nominal interest rate. One interpretation of Summers' point is that negative nominal interest rates may in some instances be desirable, with the result that mean inflation may deviate from the optimal level in order to allow for a complete response to some larger set of shocks.

In general, the greater the likelihood of a shock driving the desired nominal interest rate below zero, and the higher the loss associated with not being able to react to such a shock, the higher will be the average level of inflation that minimizes the policymaker's loss function.

To see the point, consider the simple model presented in the discussion of the output-inflation variability trade-off. Then, the restriction that $r_t \geq 0$ implies that the loss is minimized for target inflation equal to approximately $0.276\sigma_u$. That is, average inflation will be approximately one-quarter of the standard deviation of the shocks to the price level. More complex forms of the model will have similar properties. In general, the greater the likelihood of a shock driving the desired nominal interest rate below zero, and the higher the loss associated with not being able to react to such a shock, the higher will be the average level of inflation that minimizes the policymaker's loss function.

A similar result would arise when the loss function is asymmetrical. It has been argued that deflation brings potential costs that are distinct from those that come from realized inflation that is less than expected. These costs arise largely because the zero nominal interest rate floor implies that deflation beyond a certain level increases the real interest rate (ex ante and ex post), resulting in a lower steady-state capital stock.[28] This relationship suggests that realized prices below the target may be more costly than equivalent realizations above the target. This would naturally create a positive bias in the policy rule that would result in average inflation exceeding π^*.

To gauge the extent of this problem, I compute the frequency with which the ex post real interest rate has been below zero and below -1.0 percent (see table). Note that the problem is clearly most severe for the United States and France. But for other countries it is relatively modest. In fact, assuming that inflation includes an upward bias of roughly 1 percentage point, the realized real interest rates were negative less than 20 percent of the time in all countries except the United States.

Interest Rate Smoothing

Another important issue for central bank policymakers concerns the desirability of smoothing the changes in the policy instrument. There are two issues here. First is the question whether, following a shock, the optimal response is to have interest rates move immediately up (or down) and then return smoothly down (or up) to the steady-state level, always moving in the same direction following the initial jump. Second, if policymakers intend to change interest rates by some amount, should the entire change occur all at once?

The policy reaction function immediately yields the answer to the first question. Here the presumption must be that $\phi(L)$ is not monotone. That is, it does not imply movements in which interest rates jump initially and then return to the initial level, always moving in the same direction. To see this, consider Chart 4, which plots the optimal reaction of interest rates to an innovation in the aggregate price level implied by the impulse response functions plotted in Chart 1 (for the case where $h = 36$ and $\alpha = 1$).[29] The path is hump-shaped. That is, the optimal response to an innovation is to raise interest rates immediately, continue to raise them gradually, and then lower them slowly. This pattern could be further exaggerated if the loss function included an explicit cost to changing interest rates—a term of the form $k(r_t - r_{t-1})^2$.

The second question is more difficult. If the central bank were to decide that the interest rate should be increased by 100 basis points, should the change be in one large jump or in a series of smaller ones? If policy were sufficiently transparent that everyone knew that the interest rate would ultimately rise 100 basis points, so that the changes would be perfectly anticipated, then it is difficult to see why a series of smaller changes would be

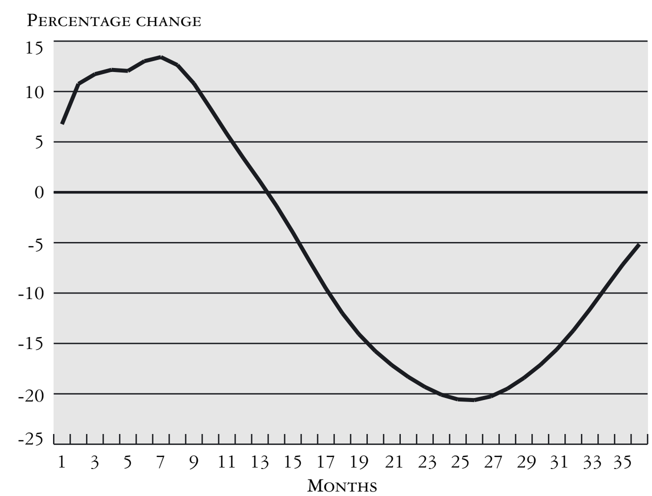
Frequency of Negative Ex Post Real Interest Rate

Country	Date of Initial Observation	Percentage of Observations That Are Greater Than Zero	Percentage of Observations That Are Greater Than -1.0
France	Jan. 1970	68	82
Japan	Nov. 1978	82	96
Germany	Jan. 1970	94	99
Italy	Nov. 1979	94	95
United Kingdom	June 1974	77	80
United States	Jan. 1970	69	78

Notes: The interest rate is the rate on three-month Treasury securities, or equivalent. Data are monthly.

Chart 4

Interest Rate Path Following a Shock



Source: Cecchetti (1996).

preferred over a single one. But often, I suspect, this question is asked with a different intention. In fact, policymakers will start to change interest rates without really knowing what the final results are likely to be. That is, uncertainty about the likely impact of the policy action on the objective will prompt policymakers to make gradual moves so that they can monitor the results—a strategy that may help improve the precision of policymaking.

Should Central Bankers Follow Rules?

The entire discussion thus far has been directed at the construction of a rule for central bank policy. But what is our motivation for constructing a set of systematic responses to external events? There are two important reasons to support the adoption of rules by the central bank. The first is the well-known finding that, when policymaking is based on pure discretion rather than rules, the dynamic inconsistency problem leads to high steady inflation. The second reason concerns the importance of policy transparency.

Over fifteen years ago, Barro and Gordon (1983) noted that if a policymaker cannot credibly commit to a zero inflation policy, then even if the policymaker announces that inflation will be zero and all private decisions are based on the assumption that inflation will in fact be zero, it is in the policymaker's interest to renege and induce inflation of some positive amount. The reason for this is that at zero inflation, the value of the increase in output obtained from fooling private agents and creating a transitory increase in output (along a Phillips or Lucas supply curve) more than offsets the cost of the higher inflation, and so the claim of zero inflation in the absence of commitment is not credible. In the language of optimal control, a zero inflation policy is not dynamically consistent.

Since the problem is thought to be most severe when potentially short-sighted legislators are capable of influencing central bank policy directly, the most prominent solution has been to create independent central banks. It is commonly thought, and the data confirm, that policymakers who are more independent are better able to make more credible commitments to low-inflation policy. [30]

As Alan Blinder (1997) has recently pointed out, however, there is a potential conflict between central bank independence and representative democracy. Since one of the crucial elements of a democratic society is that the powerful policymakers are accountable to the people, how can we square these two apparently disparate goals of accountability and independence?

Blinder (1997) and Bernanke and Mishkin (1997) suggest that the solution is policy transparency. They argue that if policymakers announce targets and are forced to explain their actions in relation to these preannounced goals, then there is accountability. Put another way, transparency and accountability are enhanced if the elected officials announce the loss function that the central bankers are charged with minimizing, and if the central bankers in turn demonstrate how they are accomplishing this goal. Researchers have suggested that the publication of the target paths for prices and/or output would serve this purpose. In fact, not only would policymakers become more accountable, but their policies would become more transparent. [31]

Arguments such as these have led to the implementation of explicit targeting regimes in a number of countries. Prominent among these countries are Australia, Canada, Finland, Israel, New Zealand, Spain, Sweden, and the United Kingdom.[32] Because explicit targeting regimes are transparent, they are easily understood. As a result, potential policy actions are less likely to create uncertainty and instability.

Lessons for Policy Formulation

This analysis offers a number of lessons. First, and most important, if a policymaker were to focus on inflation alone, the likely result—in the absence of fundamental changes in the structure of the economy—would be a very high level of real output variation. This finding provides strong support for the very flexible way in which policy targeting is currently carried out around the world.

Consider the example of the countries that have adopted explicit inflation targeting—Australia, Canada, Finland, Israel, New Zealand, Spain, Sweden, and the

United Kingdom, among others. The central banks in most of these countries appear to take into account short-to-medium-run real fluctuations when deciding on their policies. This approach is most evident in the banks' official statements. For example, the central banks in New Zealand, the United Kingdom, and a number of other countries announce target ranges—rather than point targets—for inflation. The Reserve Bank of Australia states that its goal is to have inflation average between 2 and 3 percent *over the business cycle*. By using this wording, the central bank retains the flexibility to stabilize in the face of short-run real shocks. Even countries with explicitly stated inflation targets behave as if they place some weight on output variability in their implicit loss function.

No country has adopted a zero inflation target, or even a range that is centered at zero. In fact, Haldane (1995, p. 8) reports that only New Zealand's target range includes zero at the lower end. This suggests that countries continue to be wary of the possibility of deflation and sensitive to the dangers inherent in bumping against the zero nominal interest rate floor.

The calculations in this essay also underscore the high degree of uncertainty attending the analysis of central bank policy rules. First, I note that the estimated responses of output and inflation to innovations in interest rate policy are extremely imprecise. In other words, policymakers are very unsure about the likely impact of their actions on their objectives. Since I am able to quantify this uncertainty, I can proceed to measure its impact on optimal policymaking. Thus, when I explicitly account for the imprecision of the econometric estimates needed to formulate a rule, I find that the optimal reaction of interest rates to external economic shocks declines by a factor of about twenty-five.

Let me conclude by emphasizing that substantial work remains to be done before we can convincingly articulate a detailed and operational rule for central bank policy. The framework I have proposed requires crucial information on which there is simply no general agreement at this date. What is most needed is a set of stable numerical estimates of the impact of policy actions on output and prices—as well as the ability to estimate the impact of exogenous shocks on the goals of policy.

ENDNOTES

This essay is a revised version of a paper prepared for the Symposium on Operations Research 1997*, held at Freidrich-Shiller University, Jena, Germany, on September 3-5, 1997. The author thanks Paul Bennett, Kenneth Kuttner, Margaret Mary McConnell, Lars Svensson, and Dorothy Sobol for comments. The views stated herein are those of the author and not necessarily those of the Federal Reserve Bank of New York or the Board of Governors of the Federal Reserve System.*

1. Here I paraphrase a comment made by Bennett McCallum at a conference on monetary policy in January 1993.

2. A naive reading of the recent work of Christiano, Eichenbaum, and Evans (1994a, 1994b) surely could lead to such an interpretation.

3. Throughout the discussion in this section, I assume that policymakers can credibly commit to whatever rule they choose. I return to this issue in my discussion of the dynamic consistency problem later in the essay.

4. In some formulations, the loss function includes an additional term in the change in the control variable. That is, changes in interest rates are assumed to be explicitly costly. Inclusion of such a term here adds very little to the analysis.

5. Specifying a process for y^* would be difficult because there is no agreement on several crucial issues. For example, should y^* have a random walk component or be a deterministic trend? Is the growth rate in y^* affected by the volatility of y ?

6. It is possible to nest these two objectives into a more general formulation. Consider a parameter δ representing the relative weight given to price-level targeting and inflation targeting. Then $p_t^* = \delta(\pi^* t) + (1-\delta)(p_{t-1} + \pi^*)$. The percentage of the variance in p explained by its random walk component will be related to δ .

7. The use of an interest rate is not necessary. The control variable could be any quantity that is directly governed by the central bank. For example, the monetary base or some measure of reserves could be used as the control.

8. More specifically, this is a mean zero n-variate stochastic process with finite second moments.

9. Equation 4 is the vector moving-average form. The more common vector autoregression (VAR) is equivalent.

10. This point is emphasized, for example, in Cecchetti (1995).

11. The linear-quadratic structure of the problem described here will give rise to a linear policy rule. In most cases, however, the problem would be structured differently, and the resulting rule would be more complex. For example, if the loss function were nonlinear, or there were some additional constraints on the policymaker's behavior not considered here, then the policy rule would be nonlinear as well.

12. Svensson (1996a) compares inflation and price-level targeting, arguing that one yields better performance than the other under certain economic conditions. Such an exercise relies on a particular view of the costs of inflation that is not explicitly embodied in the loss function (equation 1).

13. Ball (1997) takes a different approach, examining how the adoption of ad hoc rules that are not derived directly from the loss function will affect the loss. For example, after determining the minimized value of the loss L , he then asks how close one can get by adopting a set of arbitrary rules that do not arise from the optimal control problem itself.

14. Svensson (1997) notes that if $\alpha \neq 1$, so that weight is given to output variability in the loss function, the result would be a form of inflation forecast targeting in which the path of the forecast moves gradually back to the optimal level.

15. See the discussion of policy transparency below.

16. For a recent discussion of M2 targeting, see Feldstein and Stock (1994).

17. The procedures of the Deutsche Bundesbank reflect a different view of intermediate targets. As Mishkin and Posen (1997) note, since 1988 the German central bank has targeted growth in M3 in the belief that the demand for German M3 is stable.

18. The methods used to produce these results are described in detail in that earlier work. Briefly, I estimate a four-variable VAR including aggregate prices, commodity prices, industrial production, and the federal funds rate using monthly data from January 1984 to November 1995. Central bank policy innovations are identified and used to estimate the impulse response functions under the assumption that no variable other than policy itself responds to policy shocks immediately.

19. This point is related to the discussion of model uncertainty below.

20. The fact that the coefficients on d and s in equations 7 and 8 are all set to one is a simplification of no consequence for the main point I wish to make. Setting the variance of s to one simply means that the variance of d should be interpreted as the variance of d relative to the variance of s .

21. Cecchetti (1996) considers a substantially more complex case with the same results.

22. There are a number of ways to compute these standard errors. The simplest, called the δ-method, is to note that $\hat{\gamma}$ is a nonlinear function of the estimated parameters of the VAR. A first-order approximation of this function, together with the estimated covariance matrix of the VAR coefficients, can be utilized to compute a variance estimate for γ.

23. It is possible to obtain much more (apparent) precision by computing the average over a longer sample period. For example, using data from 1967 to 1995, the estimate of γ is -2.38, with a standard error of 1.83. But it seems absurd to argue that the reaction function embedded in the VAR is the same now as it was before 1980. This argues strongly for focusing on the estimate from the post-1984 sample.

24. Because of lags in the data, there will also be uncertainty about the current state of the economy. This type of uncertainty has an impact similar to that of parameter uncertainty considered below.

25. Blinder (1997) notes that in a multivariate model, things are not so simple, and the size and sign of covariances will determine whether policymakers exhibit more cautious or less cautious behavior.

26. A simple possibility would be to multiply the estimated covariance matrix of the estimated $A(L)$ by a positive constant.

27. The problem of inflation bias is also relevant here because measured inflation may systematically exceed true inflation. For example, Shapiro and Wilcox (1996) argue that the U.S. consumer price index may overstate inflation by 1 percentage point on average. Such a conclusion suggests that even if π^* is zero, the central bank should attempt to keep measured CPI inflation above zero.

28. See the discussion in Cecchetti (1997).

29. See Cecchetti (1996) for details on this computation.

30. Alesina and Summers (1993) establish this empirically and raise the additional possibility that countries with independent central banks not only have lower steady inflation, but also have less variable output and higher growth. Cukierman et al. (1993) also investigate the impact of central bank independence on the growth rate of output.

31. Mishkin and Posen (1997) argue that policy transparency and explicit targeting were important factors in the granting of operational independence to the Bank of England.

32. See Haldane (1995) for a discussion.

REFERENCES

Akerlof, George A., William T. Dickens, and George L. Perry. 1996. "The Macroeconomics of Low Inflation." BROOKINGS PAPERS ON ECONOMIC ACTIVITY, no. 1: 1-59.

Alesina, Alberto, and Lawrence H. Summers. 1993. "Central Bank Independence and Macroeconomic Performance: Some Comparative Evidence." JOURNAL OF MONEY, CREDIT, AND BANKING 25 (May): 151-62.

Ball, Laurence M. 1997. "Efficient Monetary Policy Rules." NBER Working Paper no. 5952, March.

Barro, Robert J., and David Gordon. 1983. "A Positive Theory of Monetary Policy in a Natural Rate Model." JOURNAL OF POLITICAL ECONOMY 91 (August): 589-610.

Bernanke, Ben S., and Frederic S. Mishkin. 1997. "Inflation Targeting: A New Framework for Monetary Policy?" NBER Working Paper no. 5893, January.

Blinder, Alan S. 1997. "What Central Bankers Learn from Academics—and Vice Versa." Distinguished Lecture on Economics in Government. JOURNAL OF ECONOMIC PERSPECTIVES 11 (spring): 3-19.

Brainard, William C. 1967. "Uncertainty and the Effectiveness of Policy." AMERICAN ECONOMIC REVIEW 57 (May): 411-25.

Cecchetti, Stephen G. 1995. "Inflation Indicators and Inflation Policy." In B. Bernanke and J. Rotemberg, eds., NBER MACROECONOMICS ANNUAL, 189-219. Cambridge: MIT Press.

————. 1996. "Practical Issues in Monetary Policy Targeting." Federal Reserve Bank of Cleveland ECONOMIC REVIEW 32, no. 1: 2-15.

————. 1997. "Understanding the Great Depression: Lessons for Current Policy." NBER Working Paper no. 6015, April.

Christiano, Lawrence J., Martin Eichenbaum, and Charles L. Evans. 1994a. "The Effects of Monetary Policy Shocks: Evidence from the Flow of Funds." Federal Reserve Bank of Chicago Working Paper no. 94-2, March.

————. 1994b. "Identification and the Effects of Monetary Policy Shocks." Federal Reserve Bank of Chicago Working Paper no. 94-7, May.

Cukierman, Alex, Pantelis Kalaitzidakis, Lawrence H. Summers, and Steven B. Webb. 1993. "Central Bank Independence, Growth, Investment, and Real Rates." CARNEGIE-ROCHESTER CONFERENCE SERIES ON PUBLIC POLICY 29 (December): 95-140.

Feldstein, Martin. 1996. "The Costs and Benefits of Going from Low Inflation to Price Stability." In C. Romer and D. Romer, eds., REDUCING INFLATION, 123-66. Chicago: University of Chicago Press.

Feldstein, Martin, and James H. Stock. 1994. "The Use of Monetary Aggregates to Target Nominal GDP." In N. Gregory Mankiw, ed., MONETARY POLICY, 7-70. Chicago: University of Chicago Press for NBER.

Groshen, Erica, and Mark Schweitzer. 1997. "Identifying Inflation's Grease and Sand Effects in the Labor Market." NBER Working Paper no. 6061, June.

Haldane, Andrew G. 1995. "Introduction." In A. G. Haldane, ed., TARGETING INFLATION, 1-12. Bank of England, November.

Lucas, Robert E., Jr. 1976. "Econometric Policy Evaluation: A Critique." CARNEGIE-ROCHESTER CONFERENCE SERIES ON PUBLIC POLICY 2: 19-46.

McCallum, Bennett T. 1997. "Issues in the Design of Monetary Policy Rules." NBER Working Paper no. 6016, April.

Mishkin, Frederic S., and Adam Posen. 1997. "Inflation Targeting: Lessons from Four Countries." Federal Reserve Bank of New York ECONOMIC POLICY REVIEW 3, no. 3: 9-110.

Shapiro, Matthew D., and David W. Wilcox. 1996. "Mismeasurement in the Consumer Price Index: An Evaluation." NBER Working Paper no. 5590, May.

Summers, Lawrence H. 1991. "How Should Long-Term Monetary Policy Be Determined?" JOURNAL OF MONEY, CREDIT, AND BANKING 23, no. 3, pt. 2 (August): 625-31.

Svensson, Lars E. O. 1996a. "Price Level Targeting vs. Inflation Targeting: A Free Lunch?" NBER Working Paper no. 5719, August.

————. 1996b. "Inflation Forecast Targeting: Implementing and Monitoring Inflation Targets." NBER Working Paper no. 5797, October.

————. 1997. "Inflation Targeting: Some Extensions." NBER Working Paper no. 5962, March.

The Expanding Geographic Reach of Retail Banking Markets

Lawrence J. Radecki

In the view of most policymakers and economists, competition in retail banking takes place in local markets covering a relatively small geographic area. Banks are thought to design their services and set their loan and deposit rates in response to the supply and demand conditions prevailing in a particular city, county, or metropolitan area. In keeping with this view, studies of the competitiveness of banking markets generally focus on developments at the local level: for example, researchers and regulatory agencies assessing the effects of bank mergers on competition will examine the degree to which deposits in a given metropolitan area are concentrated in a few large banks.

A reevaluation of the idea that banking markets are local may, however, be overdue. The banking industry has undergone a remarkable transformation in the past twenty years. Deregulation has removed many of the geographic restraints on bank expansion; banks are now free to establish branches nationwide or to buy banks in other parts of the country. In addition, banks are seeking to achieve greater efficiency in payment, credit, and depository services by standardizing their product offerings, centralizing their operations, and shifting decision-making responsibility from local managers to the head office.

In light of these changes, this article investigates whether larger geographic areas have replaced cities and counties as the true marketplace for banking services. A review of data collected during 1996 and 1997 reveals that many banks set uniform interest rates for both retail loans and deposits across an entire state or broad regions of a large state. If banks were still operating in distinct local markets, their retail interest rates would show substantial intercity variation.

Regression analysis of the effect of market concentration on deposit rates provides additional evidence that local markets have been absorbed into larger arenas of

Lawrence J. Radecki is an assistant vice president at the Federal Reserve Bank of New York.

competition: the significant relationship that earlier research detected between individual banks' deposit rates and measures of concentration at the local level is no longer evident, while a significant relationship does emerge at the state level. These results suggest that local markets the size of a single county or metropolitan area are no longer relevant and that state boundaries may offer a better approximation of the boundaries of retail banking markets.

We begin our investigation with a look at the events and ideas that have contributed to the conventional view that banking markets are local. A discussion of the forces that are reshaping the banking industry and undermining the concept of local markets follows. In the balance of the article, we present our statistical evidence supporting the emergence of larger retail markets.

How Banking Markets Have Conventionally Been Defined

The notion that retail banking markets are local in scope figured importantly in the Supreme Court's decision in the Philadelphia National Bank Case of 1963.[1] In ruling that the banking industry was subject to the nation's antitrust legislation, the Court determined that commercial banking was a bundle of services and that banking markets were local in coverage. Since then, the government agencies responsible for clearing mergers and acquisitions of banking organizations have followed the Court's lead by assessing competition within relatively narrow geographic areas.[2]

In measuring competition within local markets, regulators and other analysts have had to specify what is meant by "local." Most equate local markets in urban areas with the Census Bureau's metropolitan statistical areas (MSAs). For areas outside large cities, analysts often designate whole counties as separate markets.

Underlying the conventional definition of banking markets is the idea that market boundaries are determined from the demand side. In other words, the actions of households and business firms—the buyers of banking services—determine the reach of markets, not the actions of banks as the sellers of these services. Given the view that markets are determined from the demand side, the fact that households and businesses routinely rely on nearby institutions for most banking services has encouraged the perception that markets are quite small. Indeed, the majority of a bank's customers are typically drawn from a narrow area around each of its branch offices.

Nevertheless, commuting patterns suggest that urban markets, at least, should not be too narrowly construed. Because commuters can choose among banks convenient to their home or their workplace, they can readily switch institutions to obtain better quality or lower priced services. Recognizing that customers may be gained or lost in this way, banks operating in one part of a metropolitan area react to the price and service decisions of banks operating in other parts, even if their branch networks do not overlap. As a consequence, deposit and loan rates are highly correlated across institutions in the same metropolitan area. This correlation has supported the equation of local markets with entire metropolitan areas.

> *The view that geographic markets are local and determined from the buyer side was formed in the early 1960s, when unit banking—banks consisting of a single office—prevailed in seventeen states and branching was heavily restricted in most other states.*

Forces of Change

In the past two decades, the banking industry has undergone profound regulatory and structural changes that may make conventional definitions of markets obsolete. These changes have affected the business environment in which banks operate, the internal organization of bank holding companies, and the design and delivery of banking services.

DEREGULATION OF THE BANKING INDUSTRY

The view that geographic markets are local and determined from the buyer side was formed in the early 1960s, when unit banking—banks consisting of a single office—prevailed in seventeen states and branching was heavily restricted in most other states. As late as 1985, only twenty states permitted statewide branching. Since then, however, substantial deregulation has occurred. Unit banking has been abolished everywhere, and banks in all but five, less populous, states are permitted to establish branches throughout a state by merging with existing banks or entering de novo (Conference of State Bank Supervisors 1996).[3] These changes have led to tremendous growth in branch networks. U.S. banks in 1963 numbered 13,291, and they operated only 13,581 branch offices—a ratio of one to one. Since that time, the number of branches has quadrupled while the number of banks has shrunk. At year-end 1997, there were 60,320 branches of 9,143 banks, or more than six branches to every bank. This development alone suggests that markets now stretch beyond individual counties or metropolitan areas.

The relaxation of branching restrictions during the past two decades, coupled with numerous mergers and acquisitions, has led to substantial overlaps in banks' service areas. In the western region of New York State, for example, no bank operated branches in both Buffalo and Rochester, the region's main cities, in 1973. By 1978, only a small degree of overlap existed, with four banks operating branches in both cities (Federal Deposit Insurance Corporation 1973, 1978). Currently, however, twelve institutions operate in both cities, accounting for 94 percent of the combined $28.6 billion of deposits held in Buffalo and Rochester branches as of March 1997. Although the two metropolitan areas continue to be viewed as separate and distinct markets, the extensive overlap in branch operations indicates that retail banking in the two areas is essentially integrated.

REORGANIZATION OF HOLDING COMPANIES

Another factor that suggests the disappearance of local markets is the internal reorganization of bank holding companies. Until recently, the management of multistate holding companies was decentralized, with different charters governing company operations in different states. Within states, holding companies sometimes operated several banks, each bank confined to a distinct region and each posting a different schedule of rates for its deposit and loan products. In effect, some holding companies were confederations of separately chartered banks. To address the inefficiencies arising from redundant facilities or nonstandard products and services, many holding companies are now centralizing their management structure, organizing their operations along business—rather than geographic—lines, and placing most, if not all, banking activities under a single charter.

Another factor that suggests the disappearance of local markets is the internal reorganization of bank holding companies.

The consolidation of decision making at headquarters should encourage holding companies that now set different rates within a state to adopt uniform rates.[4] In some cases, intrastate rate differentials arose because holding companies operating several banks within a single state had a company policy of giving each bank's management some autonomy in setting consumer loan and deposit rates. Regional managers were allowed to set rates or the terms of loans and deposit accounts on the basis of their knowledge of, or feel for, local market conditions or customer preferences. In other cases, intrastate rate differentials arose because a recently acquired bank had not yet been fully integrated into its holding company.

At the same time that holding companies are reorganizing, they are making sizable investments in new technology, including credit scoring, twenty-four-hour telephone centers, and computer programs that form and analyze comprehensive customer databases. With this new technology in place, the main bank can offer an array of standardized retail products and services at all branches. Interest rate and product design decisions, based on customer research performed and interpreted by head office personnel, can be applied uniformly throughout the firm. The automation of retail services and customer support should discourage banks from reverting to their former practice of setting retail deposit and loan rates locally, even in the event of changes in underlying conditions such as a sustained rise in the general level of interest rates or further consolidation in the industry.[5]

Previous Studies of Geographic Banking Markets

Since the Supreme Court ruling in the Philadelphia National Bank case, many studies addressing the problem of market delineation have supported the position that markets are local. Early research reported the findings of surveys that collected detailed information on the location of branch offices used by households and firms in a particular municipality. These local surveys, conducted during the 1960s and 1970s, found that a large majority of individuals did their banking near home or the workplace and that small business firms generally did theirs near their establishments (Gelder and Budzeika 1970). Recent national surveys, such as the 1995 Survey of Consumer Finances and the 1993 National Survey of Small Business Finances, have found that a large majority of households and small business firms continue to use nearby institutions.[6] Although banks and other financial institutions are promoting electronic delivery of their services, only a fraction of survey respondents indicated that they use out-of-town banks.

A few econometric studies in the 1960s and 1970s attempted to identify banking markets by analyzing how interest rates varied across locations. The results of these studies were subsequently discounted, however, because deposit and loan rates were constrained by regulation at that time. The dismantling of Regulation Q, particularly the deregulation of savings and NOW accounts at year-end 1982, created the first good opportunity to inspect patterns in deposit rates to determine the size of geographic markets. The first large-scale study following deregulation, conducted by Keeley and Zimmerman (1985), yielded mixed evidence on the size of markets. The study showed statistically significant differences in average NOW account rates across metropolitan areas and individual counties in California during the 1983-84 period—a result that supports the existence of local markets. But in the case of savings accounts, Keeley and Zimmerman found that rate differences across California were too slight to indicate local markets. They also discovered that differences in state averages for savings accounts rates were large, which meant that although the market for savings account deposits was not local, it was not so large that it was national.

A study by Jackson (1992) bolstered the earlier findings of Keeley and Zimmerman by rejecting the hypothesis of a national market for both NOW accounts and savings accounts. Nevertheless, Jackson could not reject the hypothesis of a national market for six-month time deposits. Rather than perform a static comparison, as Keeley and Zimmerman had done, Jackson used time series data for individual banks over the 1983-85 period to estimate the speed with which banks adjusted retail deposit rates following changes in the Treasury bill rate. The speeds of adjustment across cities were not sufficiently similar to indicate a national market for NOW acccounts and savings accounts.

Approaching the problem from a different angle, other researchers have examined the relationship between local deposit concentration—that is, the degree to which deposits in a particular locality are concentrated in a few banks—and variations in loan and deposit rates across localities. A finding that the relationship is statistically significant provides support for the notion that markets are local.

Berger and Hannan (1989) established that measures of concentration were linked to rate differences across MSAs in the era of deregulated deposit rates. Using data for the 1983-85 period, they showed that higher degrees of local concentration were correlated with lower rates on money market savings accounts. More specifically, their analysis concluded that the savings account rate tended to run 2 basis points lower for every increase of 3 percentage points in the three-firm concentration ratio (the combined deposit share of the three largest competitors). Later studies have generally either confirmed and refined the Berger-Hannan study or extended the analysis to home mortgages and small business loans.[7]

Why a New Study of Market Size Is Warranted

Studies that have examined interest rate patterns to establish the geographic dimensions of banking markets have generally found that retail deposit or loan markets are not national. These results are often said to support the position that markets are very small and local. Nevertheless, while the hypothesis of a national market has often been rejected, a huge middle ground lies between a unified nationwide market and hundreds of markets no larger than a single county or metropolitan area. To establish the relevance of local markets, researchers need to look at data from abutting or nearby locations rather than data from cities scattered around the country.

The studies that have shown a link between deposit concentration in MSAs and differences in deposit and loan rates across cities also have important limitations. Their findings are consistent with markets that are local, but their results could also have been obtained if markets are quite a bit larger than local areas. As long as concentration in the true market area, which could encompass adjoining MSAs, is correlated with concentration in the local area, a relationship with interest rate variables would be found in the data. This means that the size of markets implied by deposit and loan rate data is still an open question.

The inconclusiveness of the existing evidence underscores the need to revisit the issue of market size. Also prompting such a reevaluation is the fact that the interest rate information used in earlier research may now be outdated. Most of the studies reviewed in the previous section relied on the findings of an annual nationwide survey of the rates and fees of retail deposit accounts in the 1983-87 period. As we have seen, banks since that time have been expanding the size and reach of their branch office networks, a development that could lead to wider geographic markets.

Interestingly, some aspects of the earlier studies hint at the possibility of wider markets in the wake of branching deregulation. First, institutions operating in a state that had unit banking or limited branching status at the time make up a sizable portion of the samples used. Neumark and Sharpe (1992) reported that one-fifth of their observations came from unit banking states and another third came from limited branching states. Second,

To establish the relevance of local markets, researchers need to look at data from abutting or nearby locations rather than data from cities scattered around the country.

some regression equations included variables that identified institutions located in unit banking or limited branching states. The estimated coefficients for location variables indicated that branching restrictions affected rate-setting behavior (Sharpe 1997). Finally, research by Hannan (1991b, 1997) showed that over the 1983-93 period, the effects of local concentration on deposit and loan rates were diminishing as branching restrictions were relaxed.

In the next two sections, we address the weaknesses in earlier research as we explore the contours of retail banking markets. First, we examine consumer deposit and loan rate data collected across cities in the same state during March 1997 to determine whether the patterns observed are consistent with the existence of local markets. If banks operate in narrowly confined markets, they should be varying retail interest rates in response to local demand and supply conditions, and intracity differences in a bank's rate schedule ought to be observed. If banks operate in broad markets, they should be setting uniform rates over regions that are wider than metropolitan areas. Uniform interest rates across an entire state would provide reasonably persuasive evidence that retail banking markets are not local.

Next we examine data collected in a 1996 survey to determine whether local concentration continues to tilt deposit rates to a bank's advantage. Uniform deposit rates over broad areas spanning several cities and the intervening regions suggest that this is no longer the case. To investigate the relationship between concentration and deposit rates more thoroughly, we use current data to reestimate some regressions specified in earlier research.

Intrastate Deposit and Loan Rate Patterns

The consumer deposit and loan data used in this section were collected by the Bank Rate Monitor, Inc., a service that provides retail pricing information for the industry.[8] The Bank Rate Monitor compiles rate information from banks in all fifty states. Although its survey tends to include only the single largest city in less populous states, it typically covers several cities in more populous states. In addition, the Bank Rate Monitor usually contacts each of the major banks at their branch offices in at least a few cities in the more populous states. The information collected on individual banks at multiple locations in the same state allows us to probe the geographic reach of markets. Here we examine six large states: New York, Michigan, Texas, California, Pennsylvania, and Florida.[9] Collectively, these states contain about 40 percent of the U.S. population.

The Bank Rate Monitor data offer a real advantage by providing rate information city by city, in contrast to previously used data sets that drew rate information only from banks' head offices. The survey does not, however, produce an ideal data set to explore the size of markets. First, only five to eight cities are surveyed in some large states. While this level of coverage may be more than adequate to meet the information needs of the survey's primary users, the performance of statistical tests requires that more cities within each state be included. Second, there are occasional gaps in coverage. The major banks in a state are not always shown to report a loan and deposit rate schedule for branches in every city included in the survey, although data on the amount of branch deposits indicate that these banks have a significant presence in some cities for which rate information is missing. In some cases, we obtained the missing information by contacting the bank directly. As a result, the data set appears to be sufficient to get a clear reading on the minimal size of markets.

Patterns in New York and Other Large States

In New York State, the Federal Reserve Bank of New York has delineated fifteen local markets that coincide roughly with metropolitan areas as defined by the Census Bureau.[10] The Bank Rate Monitor collects consumer rate information

The {Bank Rate Monitor} survey's findings show that several banks currently post uniform rate schedules for savings accounts, retail time deposits, auto loans, and home equity lines of credit across New York State.

in five local markets: Buffalo, Rochester, Syracuse, Albany, and New York City. The survey's findings show that several banks currently post uniform rate schedules for savings accounts, retail time deposits, auto loans, and home equity lines of credit across New York State (Table 1).[11] Key Bank sets identical rates for all five cities. Chase Manhattan Bank's rates, while differing from Key Bank's, are also uniform across these same cities. (It is very important to note, however, that because banks engage heavily in product differentiation through office locations and level of service, rates do not converge across competitors in the same market.) Marine Midland Bank and Fleet Bank post rates that differ from their competitors' rates but are uniform across Buffalo, Rochester, Syracuse, and Albany, a span of 294 miles. Unlike Key Bank and Chase Manhattan Bank, Marine Midland Bank and Fleet Bank, N.A., set different rates for downstate New York. The rate differentials between the banks owned by the Fleet Financial Group reflect the division of its New York State business into upstate and downstate regions and the operation of two separately chartered banks, Fleet Bank (chartered in New York) and Fleet Bank, N.A. (chartered in New Jersey). The agreement reached by Fleet Financial Group in its acquisition of National Westminster USA explains its decision to operate under two charters.

A pattern of uniform rates across an entire state is not unique to New York. Several banks in Michigan, Texas, and California post uniform rates statewide. Deposit and loan rates for a few banks are shown for the largest cities in these states in Tables 2 through 4.[12] The practice of uniform pricing, however, goes beyond the banks and cities appearing in the tables. The survey contacted ten Texas banks at both their Dallas and Houston branch offices, although only four banks are shown in Table 3. These ten jointly hold 76 percent and 70 percent of total deposits in Dallas and Houston, respectively. All ten post identical deposit and loan rates in the two cities. Uniform pricing also applies to branches of these banks in either El Paso or McCallen. The survey contacted nine California banks at their branches in both San Francisco and Los Angeles, where the banks jointly hold 65 percent and 63 percent of total metropolitan area deposits, respectively. All nine post identical rates in the two cities. Some were also contacted at branches in Bakersfield, Fresno, Modesto, or Stockton; uniform pricing was found to apply to these branches as well.

The major banks in Pennsylvania and Florida do not set uniform rates statewide, but their rates are uniform over extensive areas, spanning several local markets as currently defined (Tables 5 and 6). The patterns in these two states may not provide unqualified support of state-level markets, but they strongly contradict the use of small local markets for the analysis of competition.[13]

While it is common for banks to set uniform rates at all of their branches within a particular state, rates usually differ among branches operated by the same bank or holding company but located in different states. The banks owned

Table 1
DEPOSIT AND LOAN RATES AT SELECTED BANKS: NEW YORK STATE

Bank	Cities[a]	Money Market Deposit Account	Six-Month Time Deposit	One-Year Time Deposit	Auto Loan	Home Equity Line of Credit
Key	All five	3.01	4.25	5.75	9.25	8.25
Chase Manhattan	All five	2.79	4.65	4.71	8.95	8.25
Fleet, N.A., and Fleet	All four upstate cities	2.32	4.34	4.55	9.25	10.00
	New York City	2.27	4.29	4.39	9.25	10.00
Marine Midland[b]	All four upstate cities	2.79	5.10	5.48	10.75	9.50
	New York City	2.73	4.71	5.14	9.25	9.50
M&T Bank and East New York Savings Bank[c]	Buffalo, Rochester, New York City	2.28	5.00	5.50	9.95	8.25
First Federal Savings and Loan of Rochester[d]	Buffalo, Rochester, Syracuse, New York City	2.55	5.50	4.74	9.75	6.49

Source: Bank Rate Monitor, Inc.

[a] The five cities are the four upstate cities of Buffalo, Rochester, Syracuse, and Albany, plus New York City.

[b] Marine Midland sets rates for Nassau and Suffolk County branches that differ from those shown for New York City. According to RateGram/RateFax (reported in *Newsday*), the rate on savings accounts at the Nassau and Suffolk branches is higher than the corresponding rate at the New York City branches, while the rates on time deposits are lower.

[c] First Empire Bank Corporation owns both M&T Bank and the East New York Savings Bank but operates in the New York City area primarily through the East New York Savings Bank. The rates at the East New York Savings Bank are the same as those at M&T Bank's upstate branches. First Empire has also recently opened two supermarket branches of M&T Bank in suburban Long Island. Deposit rates at these branches are higher than the rates at the East New York Savings Bank or at M&T's upstate branches.

[d] First Federal Savings and Loan of Rochester has been acquired by HSBC Holdings, the parent of Marine Midland.

by Fleet Financial Group, for example, set uniform rates within Maine, New Hampshire, Massachusetts, Rhode Island, Connecticut, and upstate New York, but they do not set exactly the same rates in any two states (Table 7). The magnitude of these interstate rate differentials may be large enough to indicate separate markets at this time. Nevertheless, rate differentials such as these may fade away as banks take full advantage of the Riegle-Neal Interstate Banking and Branch Deregulation Efficiency Act, implemented on June 1, 1997, and as holding companies consolidate their operations into a single bank.

WHY A BANK'S RATES ACROSS LOCATIONS MIGHT CONVERGE

In principle, either the demand or the supply side of a market could be the source of pressure on a bank's interest rates in different locations to converge. But national surveys of households and small businesses find limited acceptance of

Table 2
DEPOSIT AND LOAN RATES AT SELECTED BANKS: MICHIGAN

Bank	Cities[a]	Money Market Deposit Account	Six-Month Time Deposit	One-Year Time Deposit	Auto Loan	Home Equity Line of Credit
Comerica	All five	2.30	4.60	5.10	9.25	10.25
First of America	All five	3.00	4.24	4.40	8.75	10.25
Standard Federal Savings and Loan	Detroit, Kalamazoo, Saginaw	3.25	5.00	6.00	9.00	10.25

Source: Bank Rate Monitor, Inc.

[a] The five cities are Detroit, Kalamazoo, Grand Rapids, Lansing, and Saginaw.

Table 3
DEPOSIT AND LOAN RATES AT SELECTED BANKS: TEXAS

Bank	Cities[a]	Money Market Deposit Account	Six-Month Time Deposit	One-Year Time Deposit	Auto Loan	Home Equity Line of Credit[b]
Bank One	All four	2.78	4.70	4.90	8.99	—
Bank of America	All four	3.05	4.39	4.65	13.50	—
NationsBank	All four	2.05	4.64	4.64	9.50	—
Texas Commerce	All four	2.12	4.28	4.65	9.50	—

Source: Bank Rate Monitor, Inc.

[a] The four cities are Austin, Dallas, Houston, and San Antonio.

[b] At the time of the survey, home equity lines of credit were prohibited in Texas.

Table 4
DEPOSIT AND LOAN RATES AT SELECTED BANKS: CALIFORNIA

Bank	Cities[a]	Money Market Deposit Account	Six-Month Time Deposit	One-Year Time Deposit	Auto Loan	Home Equity Line of Credit
Bank of America	All four	2.43	4.86	5.13	8.75	8.79
Wells Fargo	All four	2.38	4.87	5.15	N.R.[b]	8.92
Great Western	All four	2.50	5.35	5.50	10.75	9.24
Home Savings	All four	2.45	5.03	5.75	10.25	6.00

Source: Bank Rate Monitor, Inc.

[a] The four cities are San Francisco, Sacramento, Los Angeles, and San Diego.

[b] Not reported.

electronic banking and a strong preference for using nearby branches. Unless the responses to the survey questions are misleading or the overall findings are being misinterpreted, the surveys imply that pressure for convergence is not coming primarily from the demand, or buyer's, side.

The contrary view—that the supply side of the

Table 5
DEPOSIT AND LOAN RATES AT SELECTED BANKS: PENNSYLVANIA

Bank	Cities[a]	Money Market Deposit Account	Six-Month Time Deposit	One-Year Time Deposit	Auto Loan	Home Equity Line of Credit
CoreStates	Philadelphia	1.90	3.10	3.50	8.99	8.75
	Allentown-Bethlehem, Scranton, Harrisburg	2.00	3.50	4.00	8.00	8.75
First Union	Philadelphia, Allentown-Bethlehem, Scranton	1.00	4.00	4.25	9.49	5.75
Mellon	Philadelphia, Scranton	2.00	2.75	3.25	9.49	9.50 (9.40 in SCR)
	Harrisburg, Pittsburgh	2.02	4.25	4.65	10.50	8.99
PNC	Philadelphia	2.00	4.26	4.75	9.00	9.75
	Allentown-Bethlehem, Scranton, Pittsburgh	2.49	4.30	4.75	9.25	6.99
	Harrisburg	2.19	4.52	4.91	9.50	9.50

Source: Bank Rate Monitor, Inc.

[a]The five cities are Philadelphia, Allentown-Bethlehem, Scranton (SCR), Harrisburg, and Pittsburgh.

Table 6
DEPOSIT AND LOAN RATES AT SELECTED BANKS: FLORIDA

Bank	Cities[a]	Money Market Deposit Account	Six-Month Time Deposit	One-Year Time Deposit	Auto Loan	Home Equity Line of Credit
Barnett	Jacksonville	2.15	4.55	4.85	9.50	10.25
	Daytona Beach, Lakeland, Orlando, Melbourne	2.15	4.55	4.85	10.50	8.49
	Tampa	1.75	4.55	4.85	10.50	8.49
	Sarasota	1.75	4.55	5.00	9.50	8.49
	West Palm Beach	2.15	4.55	4.85	10.50	11.75
	Miami	2.15	4.55	4.85	10.50	8.49
First Union	Jacksonville	1.90	4.00	4.25	9.33	N.R.[b]
	Daytona Beach, Lakeland, Orlando, Melbourne	2.00	4.10	4.35	9.33	10.25
	Tampa	1.90	3.85	4.20	9.33	10.25
	Sarasota	2.00	3.85	4.20	9.33	10.25
	West Palm Beach	1.90	3.90	4.20	9.33	N.R.[b]
	Miami	1.90	4.00	4.25	9.33	10.25
NationsBank	All nine	1.01	4.15	4.60	10.00 (9.50 in WPB)	10.25
SunTrust	Jacksonville	2.20	4.81	5.00	8.50	10.25
	Daytona Beach	2.00	3.90	4.75	9.05	10.25
	Lakeland	2.00	4.75	4.95	10.35	10.25
	Orlando	2.00	4.75	4.90	8.50	10.25
	Melbourne	2.00	3.90	4.75	9.69	10.25
	Tampa, Sarasota	2.00	4.55	4.86	8.50	10.25
	West Palm Beach	2.00	4.40	4.60	8.75	7.25
	Miami	2.00	4.30	5.20	8.50	7.25

Source: Bank Rate Monitor, Inc.

[a]The nine cities are Jacksonville, Daytona Beach, Lakeland, Orlando, Melbourne, Tampa, Sarasota, West Palm Beach (WPB), and Miami.

[b]Not reported.

Table 7
DEPOSIT AND LOAN RATES ACROSS STATES: FLEET FINANCIAL GROUP

State	Money Market Deposit Account	Six-Month Time Deposit	One-Year Time Deposit	Auto Loan	Home Equity Line of Credit
Maine	2.02	3.82	4.03	9.25	10.00
New Hampshire	2.32	4.34	4.45	9.25	10.00
Massachusetts	2.17	4.18	4.45	9.25	9.75
Rhode Island	1.61	4.08	4.34	9.25	10.00
Connecticut	2.02	4.18	4.39	8.75	9.75
Upstate New York	2.32	4.34	4.55	9.25	10.00

Source: Bank Rate Monitor, Inc.

market is the source of pressure—reflects the changes that are being made in the management and operations of banks. Uniform interest rates might emerge because banks have centralized their operations and decision making at headquarters, adopted technology that diminishes the value of information collected at the branch or regional office level, or produced research showing that regional pricing does not enhance profitability. Any of these developments alone or in combination could lead a bank to regard a deposit or loan booked at one branch as a very close substitute for a comparable deposit or loan booked at another office location. Uniform rates would then come about because banks would react to a greater than expected volume of deposits taken or loans made in one part of a state by simply accepting the additional business. Banks would be less likely to respond by raising loan rates or dropping deposit rates in one location relative to rates in other cities, although at some point they might adjust a deposit or loan rate (or other terms of the deposit or loan) across the board if the total volume of that product was not meeting expectations.

A much less persuasive supply-side explanation takes into account administrative costs. Interest rates might tend to converge if administrative costs were rising so that banks could not derive any advantage—in terms of increased interest revenue or decreased interest expense—from varying their deposit and loan rates regionally. But with the trend toward greater computerization of retail operations and sharply declining prices for computer equipment, one would expect administrative costs to be falling, not rising. Therefore, administrative costs cannot readily explain the trend toward uniform retail deposit and loan rates.

HOW THE RELATIONSHIP BETWEEN CONCENTRATION AND DEPOSIT RATES IS CHANGING

Several studies using data from the mid-1980s showed that higher local concentration affected both the level of deposit rates and their speed of adjustment following changes in interest rates determined in the national money market. The uniform rates now seen over all or large parts of a bank's branch network suggest that these effects have disappeared in the wake of branching deregulation and the creation of extensive office networks. For example, the Buffalo area is characterized by higher concentration than neighboring Rochester, as measured by either the Herfindahl-Hirschman Index (HHI) or the three-firm concentration ratio.[14] Given the difference, the Berger-Hannan (1989) study would predict that money market savings rates would be 25 basis points lower in Buffalo, where banks are supposed to hold more market power, than in Rochester. But eight of the nine largest banks in Buffalo, collectively holding 94 percent of the area's deposits, set the same rate in their branches there as in their Rochester branches. Thus, savings account rates in western New York State do not appear to be influenced by local concentration. A comparison of five cities in New York State reveals that weighted and unweighted average savings account rates are similar across cities and there is no correlation between average rates and local concentration (Table 8).

In general, the breakdown of the relationship between local concentration and deposit and loan rates should occur everywhere rate uniformity is observed over a large region or an entire state. In Florida, Jacksonville is more concentrated than Miami; the three-firm concentration levels are 76 percent and 42 percent in the two cities, respectively. Three of the four banks shown in Table 6 post the same money market savings rate in the two cities, which are located at opposite ends of the state. The exception is the third largest bank in the Jacksonville area; the rate it posts in Jacksonville is 20 basis points higher than the corresponding rate in Miami, a reversal of what the concentration levels would lead one to expect. In Texas and California, the weighted and unweighted average rates are again similar across cities and they bear no relationship to local concentration (Table 8).

Table 8
AVERAGE SAVINGS ACCOUNT RATES ACROSS CITIES IN THREE STATES
Percent, Except As Noted

	New York State				
	Albany	Buffalo	New York	Rochester	Syracuse
Unweighted average	2.76	2.76	2.52	2.80	2.65
Weighted average	2.75	2.65	2.58	2.60	2.81
Banks sampled (number)	11	9	22	11	7
Combined deposit share	82	97	69	85	75
Three-firm concentration ratio	61	69	33	38	53
HHI (points)	1458	1899	748	992	1573

	Texas			
	Austin	Dallas	Houston	San Antonio
Unweighted average	2.69	2.85	2.79	2.74
Weighted average	2.46	2.53	2.49	2.72
Banks sampled (number)	6	13	14	10
Combined deposit share	51	80	76	75
Three-firm concentration ratio	41	49	41	49
HHI (points)	912	1396	890	1064

	California			
	Los Angeles	Sacramento	San Diego	San Francisco
Unweighted average	2.30	2.45	2.30	2.31
Weighted average	2.38	2.45	2.36	2.36
Banks sampled (number)	10	9	7	11
Combined deposit share	66	71	76	68
Three-firm concentration ratio	41	51	52	55
HHI (points)	900	1437	1222	1945

Sources: SNL Securities; Bank Rate Monitor, Inc.

Notes: Weights in average rates are determined by a bank's total domestic deposits. In calculations of the Herfindahl-Hirschman Index, 50 percent weighting is given to the deposits of thrifts.

The effect of statewide branch networks should also change competitive conditions in MSAs that are not headed up by large banks. Small cities in which a community bank has the leading deposit share might seem to be more susceptible to the exercise of market power than metropolitan areas with populations greater than one million. But the presence of banks operating statewide branch networks would undermine the dominance that a community bank might have in a small city. A community bank must often compete in its home town against two or more banks operating a comparable number of branches there and posting uniform and competitive rates statewide. The ability of a community bank to wield market power in this setting, even if it is the leader in market share locally, would be tightly circumscribed. The leading community bank might set lower deposit rates or higher loan rates than its main competitors, but the reason would be the higher costs associated with product differentiation (for example, more convenient office locations or longer hours), not market power.

ESTIMATED EFFECT OF CONCENTRATION ON DEPOSIT RATES

The uniformity of several banks' deposit and loan rates across an entire state suggests that state boundaries now approximate the shape and extent of retail markets better than county lines or MSA designations. To investigate the expansion of retail markets more systematically, we use regression techniques to estimate the effect of local concentration on deposit rates. Recent data on deposit rates are drawn from the *Monthly Survey of Selected Deposits and Other Accounts*, the same source used in many of the studies reviewed earlier. The survey, conducted by the Board of Governors of the Federal Reserve System, collects information on checking and savings accounts and time deposits from 399 commercial banks and thrift institutions nationwide. Although the participants represent only 4 percent of all commercial and savings banks in the country, they operate about one-quarter of all banking offices.

For each type of account—savings, checking, and time—the respondents to the survey report the interest rate that is applicable to the largest volume of deposits.[15] That is, a bank may offer two or more types of savings accounts and may vary the interest rate and other terms of each type by location, but it will report the rate that applies to the largest dollar volume of savings account deposits.

A difficulty encountered in the analysis of this data set is formulating the appropriate treatment of a bank whose branch office network spans two or more local areas. If a bank varies deposit rates by location, the city offering the interest rate reported by the survey cannot be determined. We replicate the methodology of previous studies to ensure a close correspondence between the rate reported by the survey and the MSA to which a respondent is assigned. First, any respondent that has more than 25 percent of its deposits booked at branches outside its base of operations—the city where its head office is located and it presumably does the largest share of its business—is dropped from the sample. Second, a respondent that is retained in the sample enters the analysis only in its home city. It does not enter the analysis in any other city, even one in which it holds the largest share of local deposits. Taking these two steps increases the likelihood that a bank's response pertains to the city to which it is assigned. On the downside, however, these steps diminish the coverage of the sample markedly by filtering out many of the large participants in the survey.[16] With the expansion of branch networks during the past fifteen years, these two steps should now eliminate proportionately more survey participants than before and may undermine the reliability of the regression results.

The uniformity of several banks' deposit and loan rates across an entire state suggests that state boundaries now approximate the shape and extent of retail markets better than county lines or MSA designations.

Table 9 reports the effects of extracting a usable sample from the survey. In keeping with the practice of focusing on urban areas, established in earlier studies, we first pare the list of survey respondents by eliminating 91 rural banks. (These 91 banks—mostly small institutions—have a larger proportion of deposits at branches located in non-MSA counties than in any single MSA.) The list is pared further by eliminating another 108 banks whose operations are not concentrated geographically. These mostly large institutions operated 16,401 branches, more than two-thirds of the total number of branches covered by the survey. After all trimming is performed, the sample consists of 200 banks and retains 18 percent of the branches and 29 percent of the aggregate deposits covered by the survey. Thirty-three states (and the District of Columbia) and 91 MSAs, out of a total of 317 MSAs in the nation, are represented in the sample; in 10 of the 33 states, all banks are assigned to the same MSA. The sample provides coverage in the 91 MSAs that is less thorough than the number of survey participants and their size would suggest. About 5 percent of the aggregate number of banks in the 91 covered MSAs are included in the sample; they operated 12 percent of total branches in these MSAs.

Estimation Results

To investigate the effects of local concentration under present conditions, we use the sample just described to reestimate the regression equation specified in some earlier studies. A bank's deposit rate for a savings account, NOW account, or six-month time deposit is explained in the regression by concentration in the MSA (measured by the HHI) and some control variables: (1) the bank's total assets, to account for differences among banks related to their size; (2) the population of the MSA to which the bank is assigned, to account for differences among local areas related to their size; and (3) dummy variables for each census division, to account for regional differences in wage rates, population density, or any other relevant characteristic.

{Our} results indicate that concentration at the local level no longer matters for interest rates paid to retail depositors.

Definitions of the variables are listed in Table 10; results are presented separately for the three types of deposits in Tables 11 through 13 (column 1) and compared with results reported by Hannan using 1993 and 1985 data (columns 3 and 4). Overall, the estimated coefficients and R^2 of the regression derived from 1996 data are comparable to those derived from 1993 data, but the estimated coefficient of the concentration variable for all three types of deposits is not significant (and, contrary to expectations, it is not even negative). These results indicate that concentration at the local level no longer matters for interest rates paid to retail depositors. By contrast, the importance of concentration in the mid-1980s is indicated by the high significance of the concentration variable in the savings account equation estimated using 1985 data (t-statistic of -6.79, shown in Table 11) and confirmed in other studies using data from the same era.

We estimate some additional sets of regressions to test the sensitivity of our results to the list of control variables and the definition of the concentration variable. When the control variables are expanded to include MSA income, MSA deposit growth, a bank's share of total MSA deposits, a dummy variable for thrift institutions, and a dummy variable for limited branching states—variables used in at least one of the earlier studies—coefficient estimates and t-statistics change only marginally

Table 9
COVERAGE OF BANKS PROVIDED BY THE SURVEY AND THE SAMPLE

Banks	Number	Branches	Number	Deposits	Dollar Volume
Banks in survey	399	Branches operated by the 399 banks	22,983	Deposits held at the 22,983 branches	1.28 trillion
less		*less*		*less*	
Banks located outside MSAs	91	Branches operated by these 91 banks	1,657	Deposits at these 1,657 branches	51 billion
less		*less*		*less*	
Banks that are not concentrated geographically	108	Branches operated by these 108 banks	16,401	Deposits at these 16,401 branches	822 billion
equals		*less*		*less*	
Banks in sample	200	Branches operated by these 200 banks outside the "home" MSA	803	Deposits at these 803 branches	34 billion
		equals		*equals*	
		Branches in sample	4,122	Deposits at branches in sample	370 billion

MEMO:
SUMMARY STATISTICS FOR THE NINETY-ONE MSAS INCLUDED IN THE SAMPLE

	Percentage of All Banks	Percentage of All Branches
Included in survey[a]	7.0	38
Included in sample[a]	4.7	12
Mean value of the percentage included	6.0	11
Median value of the percentage included	5.3	7
Upper quartile	7.7	16
Lower quartile	3.1	3
Maximum	23	41
Minimum	0.8	0.24

Note: The sample is drawn from the *Monthly Survey of Selected Deposits and Other Accounts* of the Board of Governors of the Federal Reserve System.

[a]In calculations of the percentage of banks included in the survey or the sample, a bank is counted multiple times if it has offices in two or more of the ninety-one metropolitan statistical areas.

(reported in column 2 of Tables 11 through 13).[17] If we give the deposits of thrift institutions either 50 percent or 100 percent weighting in the calculation of HHI instead of zero percent weighting—a reasonable modification to make if thrifts are important or full-fledged competitors of banks for household customers—the estimated coefficient on the concentration variable in the time deposit regression turns negative; however, this coefficient is still not significant. The t-statistics are -1.41 and -1.51, respectively, for 50 percent and 100 percent weighting of thrift institution deposits. (These results are not reported in the tables.) If the three-firm concentration ratio is substituted for the HHI as the measure of market concentration, results change marginally. (Again, the results are not reported.)

Table 10
LIST OF VARIABLES USED IN REGRESSIONS

Variable	Definition or Explanation	Sample Means[a] 200-Bank Sample	316-Bank Sample	390-Bank Sample
Savings account rate	Interest rate offered on money market savings accounts	2.59	2.49	2.54
NOW account rate	Interest rate offered on interest-bearing checkable deposit accounts	1.74	1.62	1.74
Time deposit rate	Interest rate offered on retail six-month time deposits	4.67	4.57	4.63
HHI	Herfindahl-Hirschman Index of concentration			
	Zero weight assigned to thrifts	MSA 1784	State 1191	State 1134
	50 percent weight assigned to thrifts	1357	888	860
	100 percent weight assigned to thrifts	1183	747	732
Three-firm concentration ratio	Sum of three largest deposit shares			
	Zero weight assigned to thrifts	MSA 63.3	State 50.3	State 49.0
	100 percent weight assigned to thrifts	50.5	40.0	39.5
Bank's total assets	Billions of dollars	3.54	5.74	4.67
Population	Millions	MSA 2.65	State 10.24	State 9.57
Average household income in MSA	Thousands of dollars	52.5	—	—
Per capita income in state	Thousands of dollars	—	18.9	18.7
Deposit growth	Percent	MSA 2.80	State 3.37	State 3.46
Bank's share of total deposits	Percent	MSA 6.39	State 4.72	State 3.90
Thrift institution	Number of institutions	39	52	57
Limited branching state	Number of institutions in AK, GA, KY, MT, OK, and WY	13	17	27

Sources: Board of Governors of the Federal Reserve System, *Monthly Survey of Selected Deposits and Other Accounts*; SNL Branch Migration Data Base (version 6.1); Federal Deposit Insurance Corporation.

[a]Three sets of regressions are estimated using different sample sizes, corresponding to the number of observations used in local-level regressions, state-level regressions excluding rural banks, and state-level regressions including rural banks. The sample sizes reflect the number of observations used in the savings account regressions. One to three fewer observations were used in the NOW account and time deposit regressions because of missing data.

STATE-LEVEL ANALYSIS

Next we estimate regression equations comparable to those just discussed to see whether concentration at the state level influences retail deposit rates. Some variables

Table 11
THE RELATIONSHIP BETWEEN A BANK'S SAVINGS ACCOUNT DEPOSIT RATE AND LOCAL AREA CONCENTRATION

Explanatory Variables	Year in Which Survey Was Conducted: 1996 (1)	1996 (2)	1993 (Hannan 1997) (3)	1985 (Hannan 1991b) (4)
Intercept	2.35 (10.85)	2.56 (5.76)	2.62 (20.79)	7.12 (96.05)
MSA HHI (zero weight assigned to thrifts)	0.38E-4 (0.53)	0.51E-4 (0.68)	-0.46E-4 (-0.99)	-2.32E-4 (-6.79)
Bank total assets	0.22E-2 (0.43)	0.68E-2 (1.20)	-0.64E-2 (-2.25)	0.53E-2 (0.91)
MSA population	0.11E-1 (0.53)	0.99E-2 (0.44)	-0.23E-1 (-2.25)	-1.52E-2 (-1.26)
Per capita income in MSA		-0.57E-2 (-0.88)		
MSA deposit growth		-0.49E-2 (-0.50)		
Bank's share of total MSA deposits		-0.75 (-1.20)		
Thrift institution		-0.17 (1.13)		
Limited branching state		-0.32 (-1.52)		
Memo:				
Number of observations	200	200	341	330
R^2	0.061	0.091	0.074	0.124

Notes: Regional dummy variables are included in the 1993 and 1996 regressions, but the estimated coefficients are not reported. In the 1985 regression, the annual rate of business failures in the state in which a bank is located is included; the estimated coefficient for this variable is 0.12E-3 (1.26). Figures in parentheses are t-statistics.

used earlier are redefined in order to take this step: deposit concentration at the state level replaces deposit concentration at the MSA level, state population replaces MSA population, and so forth. The estimates are reported in Tables 14 through 16. The first set of state-level regressions (column 1) are estimated using almost the same sample of banks as before at the local level.[18] In this first set of regressions, the estimated coefficient on the concentration variable turns negative for all three deposit rates but is still insignificant.

The second and third sets of state-level regressions (columns 2 and 3) use a larger sample of 316 survey respondents because it is no longer necessary to match a bank with an MSA. Only small rural banks are now excluded.[19] This adjustment sharply improves the sample's coverage. With the return of 122 large banks, all but 345 branches covered by the survey are now included in the sample. In this pair of regressions, the estimated coefficient on the concentration variable has a negative sign and becomes significant in the savings account equation, but is still insignificant in the NOW account and time deposit equations. In the fourth and fifth sets of regressions (columns 4 and 5), the HHI measure is replaced by the three-firm concentration ratio. Zero weight is given to thrift institution deposits in the fourth regression, but 100 percent weight is given in the fifth regression. The estimated coefficient of the concentration variable is significant in both the savings account and NOW account equations, but still insignificant in the time deposit equation. Additional regressions are estimated (although not reported in the tables) in which 100 percent weight is given to thrift deposits in calculations of the the HHI, or extra control variables are included in the list of explanatory variables. The estimated coefficient for the

Table 12
THE RELATIONSHIP BETWEEN A BANK'S NOW ACCOUNT DEPOSIT RATE AND LOCAL AREA CONCENTRATION

Explanatory Variables	Year in Which Survey Was Conducted 1996 (1)	1996 (2)	1993 (Hannan 1997) (3)
Intercept	1.42	1.49	1.72
	(8.30)	(4.29)	(12.36)
MSA HHI	0.78E-4	0.96E-4	-0.54E-4
(zero weight assigned to thrifts)	(1.43)	(1.63)	(-1.06)
Bank total assets	-0.73E-2	-0.19E-3	-0.92E-2
	(-1.79)	(-0.42)	(-2.98)
MSA population	-0.39E-2	-0.45E-2	-0.39E-2
	(-2.43)	(-2.55)	(-3.45)
Per capita income in MSA		-0.35E-2	
		(-0.69)	
MSA deposit growth		-0.28E-2	
		(-0.37)	
Bank's share of total MSA deposits		-0.88E-2	
		(-1.77)	
Thrift institution		0.19	
		(1.64)	
Limited branching state		-0.20	
		(-1.20)	
Memo:			
Number of observations	197	197	341
R^2	0.212	0.245	0.254

Notes: Regional dummy variables are included, but the estimated coefficients are not reported. Figures in parentheses are t-statistics.

Table 13
THE RELATIONSHIP BETWEEN A BANK'S SIX-MONTH TIME DEPOSIT RATE AND LOCAL AREA CONCENTRATION

Explanatory variables	Year in Which Survey Was Conducted 1996 (1)	1996 (2)	1993 (Hannan 1997) (3)
Intercept	4.76	4.47	2.75
	(24.60)	(11.90)	(23.55)
MSA HHI	0.24E-5	0.34E-4	-0.63E-4
(zero weight assigned to thrifts)	(0.04)	(0.55)	(-1.50)
Bank total assets	-0.33E-2	0.59E-2	-0.66E-2
	(-0.72)	(1.20)	(-2.60)
MSA population	-0.99E-2	-0.34E-1	-0.14E-1
	(-0.56)	(-1.84)	(-1.46)
Per capita income in MSA		0.55E-3	
		(0.01)	
MSA deposit growth		0.49E-2	
		(0.60)	
Bank's share of total MSA deposits		-1.70E-2	
		(-3.30)	
Thrift institution		0.44	
		(3.55)	
Limited branching state		0.28	
		(1.60)	
Memo:			
Number of observations	197	197	320
R^2	0.059	0.182	0.092

Notes: Regional dummy variables are included, but the estimated coefficients are not reported. Figures in parentheses are t-statistics.

Table 14

THE RELATIONSHIP BETWEEN A BANK'S SAVINGS ACCOUNT DEPOSIT RATE AND CONCENTRATION AT THE STATE LEVEL

Explanatory Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Intercept	2.71	3.00	2.03	3.55	2.99	3.04	3.60
	(10.12)	(13.24)	(4.02)	(8.98)	(12.01)	(16.27)	(11.24)
State HHI (zero weight assigned to thrifts)	-0.14E-3	-0.22E-3	-0.33E-3			-0.24E-3	
	(-1.28)	(-2.29)	(-3.21)			(-3.00)	
State three-firm concentration ratio				-0.15E-1	-0.10E-1		-0.16E-1
(weight assigned to thrifts is shown in italics)				(-2.64)	(-1.96)		(-3.42)
				zero	*100 percent*		*zero*
Bank total assets	0.42E-4	-0.17E-2	-0.37E-2	-0.15E-2	-0.16E-2	-0.32E-2	-0.29E-2
	(0.009)	(-0.58)	(-0.98)	(-0.49)	(-0.52)	(-1.12)	(-1.01)
State population	0.13E-1	-0.44E-3	0.23E-3	-0.60E-3	-0.25E-2	-0.25E-2	-0.27E-2
	(1.60)	(-0.07)	(0.03)	(-0.09)	(-0.37)	(-0.42)	(-0.45)
Per capita income in state			0.44E-1				
			(1.92)				
State deposit growth			0.34E-1				
			(1.56)				
Bank's share of total state deposits			0.10E-1				
			(1.38)				
Thrift institution			0.32				
			(2.61)				
Limited branching state			-0.90E-1				
			(-0.49)				
Number of observations	194	316	316	316	316	390	390
R^2	0.088	0.070	0.114	0.075	0.065	0.073	0.079

Notes: Regional dummy variables are included but their estimated coefficients are not reported. Figures in parentheses are t-statistics.

Table 15

THE RELATIONSHIP BETWEEN A BANK'S NOW ACCOUNT DEPOSIT RATE AND CONCENTRATION AT THE STATE LEVEL

Explanatory Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Intercept	1.68	1.59	0.82	2.10	1.99	1.82	2.42
	(7.90)	(8.78)	(2.06)	(6.67)	(10.15)	(11.16)	(8.70)
State HHI (zero weight assigned to thrifts)	-0.36E-4	-0.36E-4	-0.58E-4			-1.08E-4	
	(-0.42)	(-0.48)	(-0.72)			(-1.54)	
State three-firm concentration ratio				-0.86E-2	-0.11E-1		-0.12E-1
(weight assigned to thrifts is shown in italics)				(-1.92)	(-2.64)		(-3.07)
				zero	*100 percent*		*zero*
Bank total assets	-0.12E-1	-0.97E-2	-0.40E-2	-0.94E-2	-0.93E-2	-0.13E-1	-0.12E-1
	(-3.31)	(-4.05)	(-1.38)	(-3.96)	(-3.91)	(-5.04)	(-4.91)
State population	-0.10E-1	-0.91E-2	-0.19E-1	-0.10E-1	-0.13E-1	-0.12E-1	-0.13E-1
	(-1.51)	(-1.72)	(-3.16)	(-1.93)	(-2.44)	(-2.28)	(-2.45)
Per capita income in state			0.37E-1				
			(2.06)				
State deposit growth			-0.70E-2				
			(-0.41)				
Bank's share of total state deposits			-0.14E-1				
			(-2.38)				
Thrift institution			0.27				
			(2.83)				
Limited branching state			-0.43E-1				
			(-0.29)				
Number of observations	192	314	314	314	314	387	387
R^2	0.181	0.141	0.215	0.151	0.160	0.195	0.210

Notes: Regional dummy variables are included but their estimated coefficients are not reported. Figures in parentheses are t-statistics.

concentration variable is almost always significant in the savings account regressions; the t-statistic is highest when zero weight is given to thrift deposits and extra control variables are included. The estimated coefficient for the concentration variable, however, is never significant at the 5 percent level in the additional NOW account and time deposit regressions.

Lastly, we estimate the regressions using an almost complete set of survey respondents, including small rural banks. For this larger sample of 390 observations, we report the results from two sets of regressions—one using the state HHI as the concentration measure and the other using the state three-firm concentration ratio—in columns 6 and 7. The estimate of the coefficient of the concentration measure is significant in both regressions for the savings account rate, and the estimate is also significant in the NOW account rate equation that uses the three-firm concentration ratio. As before, we estimate additional regressions in which 100 percent weight is given to thrift deposits or extra control variables are included. Although the results are not reported in the table, the estimated coefficient for the concentration variable is always significant in the savings account regressions; the t-statistics are in the range of -2.38 to -3.74. The estimated coefficient for the concentration variable is significant at the 10 percent level in half of the regressions explaining the NOW account rate. The t-statistics fall in the range of +0.51 to -3.13 and are

{Our} estimates indicate that an increase in concentration at the state level will have an economically meaningful effect on savings account rates.

Table 16

THE RELATIONSHIP BETWEEN A BANK'S RETAIL SIX-MONTH CERTIFICATE OF DEPOSIT RATE AND CONCENTRATION AT THE STATE LEVEL

Explanatory Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Intercept	5.23	4.79	4.45	4.81	5.02	4.71	4.64
	(20.53)	(24.08)	(10.57)	(13.97)	(23.35)	(28.00)	(16.12)
State HHI (zero weight assigned to thrifts)	-0.20E-3	-0.45E-4	-0.42E-6			-0.14E-4	
	(-1.94)	(-0.54)	(-0.01)			(-0.20)	
State three-firm concentration ratio				-0.18E-2	-0.75E-2		0.65E-3
(weight assigned to thrifts is shown in italics)				(-0.37)	(-1.69)		(0.15)
				zero	*100 percent*		*zero*
Bank total assets	-0.44E-2	-0.59E-2	0.20E-2	-0.59E-2	-0.56E-2	-0.74E-2	-0.75E-2
	(-1.01)	(-2.25)	(0.65)	(-2.25)	(-2.16)	(-2.89)	(-2.91)
State population	-0.84E-2	-0.27E-2	-0.16E-1	-0.25E-2	-0.54E-2	-0.55E-2	-0.53E-2
	(-1.08)	(-0.47)	(-2.47)	(-0.44)	(-0.91)	(-1.03)	(-0.99)
Per capita income in state			0.66E-3				
			(0.04)				
State deposit growth			0.27E-1				
			(1.49)				
Bank's share of total state deposits			-0.21E-1				
			(-3.26)				
Thrift institution			0.52				
			(5.09)				
Limited branching state			0.13				
			(0.84)				
Number of observations	193	315	315	315	315	389	389
R^2	0.079	0.071	0.187	0.071	0.079	0.076	0.076

Notes: Regional dummy variables are included, but their estimated coefficients are not reported. Figures in parentheses are t-statistics.

highest using the three-firm concentration ratio.

Our regression results provide estimates of the effect of greater concentration on savings account rates: an increase of 20 percentage points in the three-firm concentration level causes savings account rates to fall on the order of 20 to 30 basis points. The estimated effect of a substantial increase in the HHI on savings account rates is comparable: a 1000 point increase in the index causes rates to fall 25 basis points. These estimates indicate that an increase in concentration at the state level will have an economically meaningful effect on savings account rates.

CONCLUSION

For many years, analysts seeking to delineate geographic markets for retail banking services have referred to demand forces and consequently have judged banking markets to be small and local. The current practice among banks in New York and other large states, however, is to set uniform retail deposit and consumer loan rates across an entire state or large regions of a state. This pattern implies that the geographic reach of these markets is much larger than a metropolitan area. Furthermore, a shift to broader markets, determined from their supply side, is a development that is congruent with the growth of branch office networks and with the changes implemented by holding companies in both their operations and their internal organization.[20]

Estimates of the relationship between retail deposit rates and measures of market concentration provide further evidence that banking markets have expanded. Using 1996 data, this analysis finds that the statistically significant correlation that existed at the local level in the mid-1980s has disappeared. In addition, the analysis finds a significant correlation at the state level for some measures of concentration and some deposit rates. Against this background, markets now appear to be at least as large as a state, but how much larger is less clear. Our intuition tells us that markets are unlikely to be perfectly coincident with state borders. Nevertheless, state borders offer a better approximation of the territory over which banks compete for household customers than do counties or metropolitan areas.

The scope of markets may stretch beyond individual states fairly soon, however, with the advent of full interstate branching and further consolidation. The choices of households may also promote expansion of geographic markets from the demand side. Many individuals currently hold shares of mutual funds, and half of all mutual fund accounts are opened with sponsors whose marketing tools are mainly confined to the mail and telephone. Even now, some bankers comment that a sizable proportion of customers rarely, if ever, come into a branch office. If depositors are offered incentives in the form of higher yields or lower minimum balance requirements, many might be prepared to switch to an out-of-town bank, a development that would create a national market for retail banking products.

Significantly, larger retail banking markets may be more competitive than is commonly perceived. For many years, the public did not regard retail banking as a highly competitive business because branching restrictions protected local markets for depository institutions. Despite the lifting of these restrictions, it seems that few people believe that vigorous competition has broken out. This article's finding that markets are growing larger in geographical scope casts doubt on the persistent belief that competition is weak. Because the industry is less concentrated at the state and national levels than at the MSA level, competition among banks should be spirited.

ENDNOTES

The author thanks Joseph Doyle for research assistance and helpful comments throughout the preparation of this study.

1. See *United States v. Philadelphia National Bank*, 374 U.S. 321 (1963).

2. It is recognized, however, that certain products, such as all-purpose credit cards, are offered in a national setting.

For a description of current procedures for defining markets and evaluating the level of competition in these markets, see Amel (1997) and Herlihy et al. (1997).

3. The states are Arkansas, Georgia, Kentucky, Oklahoma, and Wyoming.

4. Banc One Corporation, which has operated seventeen banks and used seventeen corresponding pricing regions in Ohio, is planning to consolidate operations in the state into a single bank and to offer identical checking and savings account rates at all branches, although it will use three regions to set rates on certificates of deposit. See *Bank Rate Monitor* (1997).

5. Although this article argues that organizational and technological changes will promote uniform rates, Calem and Nakamura (1997) have developed a theoretical model, based on a Bertrand pricing game, that predicts uniform rate setting.

6. Kennickell, Starr-McCluer, and Sundén (1997) summarize the findings of the most recent household survey and Cole and Wolken (1995) the findings of the small business survey.

7. Among these studies are Hannan (1991a, 1997), Hannan and Berger (1991), Neumark and Sharpe (1992), Rhoades (1992), and Sharpe (1997).

8. The Bank Rate Monitor standardizes the information it obtains on loan rates by using the following criteria: Auto loan rates are based on a $16,000 new car loan with 10 percent down and a four-year term. Home equity line of credit rates are for an open-ended line and are based on the minimum amount that can be borrowed or the minimum credit line, whichever applies. Rates offered may be introductory.

9. Illinois and New Jersey are two large states that could not be included because the survey covers only Chicago and Newark.

10. Ten New York banking markets center on a city designated as the core of an MSA. The other five center on a city that is not part of an MSA.

11. The practice of setting uniform rates for savings and NOW accounts was observed in California as early as the mid-1980s (Keeley and Zimmerman 1985).

12. The data cover deposit and loan rates for households but not for small business firms. Nevertheless, uniform rates and fees seem to apply to these firms as well. Information from some banks indicates that a single schedule of terms and fees is set for small business checking accounts throughout a state.

13. Ohio is a large state in which regional deposit rate patterns are observed. The large holding companies have each operated multiple banks in the state but may soon consolidate them and change their rate-setting strategies (see endnote 4).

14. The HHI is defined as the sum of the squared market shares of all banking organizations operating in an area. We calculated the HHI for urban markets using branch deposit data collected June 30, 1996, and information on bank ownership as of April 21, 1997.

15. Banks are asked to supplement their responses to the survey by providing information on rate tiers and corresponding balance requirements. In the regressions, the lowest rate reported is used.

16. Control variables are added to the regression equation to account for differences among local markets and among banks. Measurement of control variables also becomes problematic for banks whose branch network spans several cities.

17. The control variables are expected to play a more important role in state-level regressions than in MSA-level regressions because MSAs are made to be fundamentally similar in their construction, while states are very different from one another.

18. Two money center banks are excluded because they have no retail operations. Delaware banks are excluded because state concentration measures are skewed by the presence of large credit card banks. A District of Columbia bank is also excluded.

19. The sample is increased first by bringing back banks that could not be matched reliably with a single MSA. Then the largest of the rural banks (those holding more than $1 billion of assets) are added because an examination of their deposit base found that a substantial proportion of their deposits were held at branches located in MSAs.

20. The level of competition in small business lending has also been evaluated for many years in the context of very local markets. A parallel trend toward broader geographic markets may also be occurring for this banking product. While active competition in the supply of small business credit is certainly a concern of policymakers, this topic is beyond the scope of the article.

REFERENCES

Amel, Dean F. 1997. "Antitrust Policy in Banking: Current Status and Future Prospects." Paper presented at the 1997 Bank Structure and Competition Conference, Federal Reserve Bank of Chicago.

Bank Rate Monitor. 1997. "Banc One Plans to Go to National Pricing." Vol. 16, no. 15, pt. 2 (April): 3, 9.

Berger, Allen N., and Timothy H. Hannan. 1989. "The Price-Concentration Relationship in Banking." REVIEW OF ECONOMICS AND STATISTICS 71, no. 2: 291-9.

Calem, Paul S., and Leonard I. Nakamura. 1997. "Branch Banking and the Geography of Bank Pricing." Unpublished paper.

Cole, Rebel A., and John D. Wolken. 1995. "Financial Services Used by Small Businesses: Evidence from the 1993 National Survey of Small Business Finances." FEDERAL RESERVE BULLETIN 81, no. 7: 629-67.

Conference of State Bank Supervisors. 1996. A PROFILE OF STATE CHARTERED BANKING.

Federal Deposit Insurance Corporation. 1973. OPERATING BRANCH OFFICES.

———. 1978. DEPOSIT DIRECTORY.

Gelder, Ralph H., and George Budzeika. 1970. "Banking Market Determination—The Case of Central Nassau County." Federal Reserve Bank of New York MONTHLY REVIEW: 258-66.

Hannan, Timothy H. 1991a. "Bank Commercial Loan Markets and the Role of Market Structure: Evidence from the Surveys of Commercial Lending." JOURNAL OF BANKING AND FINANCE 15, no. 1: 133-49.

———. 1991b. "The Functional Relationship Between Prices and Market Concentration: The Case of the Banking Industry." Board of Governors of the Federal Reserve System, Finance and Economics Discussion Series, no. 169.

———. 1997. "Market Share Inequality, the Number of Competitors, and the HHI: An Examination of Bank Pricing." REVIEW OF INDUSTRIAL ORGANIZATION 12: 23-35.

Hannan, Timothy H., and Allen N. Berger. 1991. "The Rigidity of Prices: Evidence from the Banking Industry." AMERICAN ECONOMIC REVIEW 81, no. 4: 938-45.

Herlihy, Edward D., David S. Neill, Craig M. Wasserman, Adam D. Chinn, John C. Coates IV, and Nancy M. Clark. 1997. "Financial Institutions—Mergers and Acquisitions 1996: Another Successful Round of Consolidation and Capital Management." In FIFTEENTH ANNUAL BANKING EXPANSION INSTITUTE, 319-598. Little Falls, N.J.: Glasser Legal Works.

Jackson, William E. III. 1992. "Is the Market Well Defined in Bank Merger and Acquisition Analysis?" REVIEW OF ECONOMICS AND STATISTICS 74, no. 4: 655-61.

Keeley, Michael, and Gary C. Zimmerman. 1985. "Determining Geographic Markets for Deposit Competition in Banking." Federal Reserve Bank of San Francisco ECONOMIC REVIEW, summer: 25-45.

Kennickell, Arthur B., Martha Starr-McCluer, and Annika E. Sundén. 1997. "Family Finance in the U.S.: Recent Evidence from the Survey of Consumer Finances." FEDERAL RESERVE BULLETIN 83, no. 1: 1-24.

Neumark, David, and Steven A. Sharpe. 1992. "Market Structure and the Nature of Price Rigidity: Evidence from the Market for Consumer Deposits." QUARTERLY JOURNAL OF ECONOMICS 107, no. 2: 657-80.

Rhoades, Stephen A. 1992. "Evidence on the Size of Banking Markets from Mortgage Loan Rates in Twenty Cities." Board of Governors of the Federal Reserve System Staff Study no. 162.

Sharpe, Steven A. 1997. "The Effect of Consumer Switching Costs on Prices: A Theory and Its Application to the Bank Deposit Market." REVIEW OF INDUSTRIAL ORGANIZATION 12, no. 1: 79-94.

Dealers' Hedging of Interest Rate Options in the U.S. Dollar Fixed-Income Market

John E. Kambhu

As derivatives markets have grown, the scope of financial intermediation has evolved beyond credit intermediation to cover a wide variety of risks. Financial derivatives allow dealers to intermediate the risk management needs of their customers by unbundling customer exposures and reallocating them through the derivatives markets. In this way, a customer's unwanted risks can be traded away or hedged, while other exposures are retained. For example, borrowers and lenders can separate a loan's interest rate risk from its credit risk by using an interest rate swap to pass the interest rate risk to a third party. In another example of unbundling, an option allows an investor to acquire exposure to a change in asset prices in one direction without incurring exposure to a move in asset prices in the opposite direction.

John E. Kambhu is an assistant vice president at the Federal Reserve Bank of New York.

The derivatives markets' rapid growth has been driven by a number of developments. In addition to advances in finance and computing technology, the rough balance of customer needs on the buy and sell sides of the market has contributed to this expansion. This balance allows dealers to intermediate customer demands by passing exposures from some customers to others without assuming excessive risk themselves. Without this ability to pass exposures back into the market, the markets' growth would be constrained by dealers' limited ability to absorb customers' unwanted risks.

The balance between customer needs on both sides of the market is most apparent in the swaps market, the largest of the derivatives markets, where only a small amount of residual risk remains with dealers.[1] In the over-the-counter U.S. dollar interest rate options market, however, significant residual risks are concentrated among dealers, who have sold 50 percent more options to customers than they have purchased (Table 1, top panel). This imbalance has left dealers with significant net exposure to price risk that must be hedged in the underlying fixed-income markets.

Until now, the scale of hedging across all dealers in the over-the-counter interest rate options market has not been studied in the literature. The concentration of sold options among dealers, however, makes it an ideal place to explore how dealers' hedging of options affects underlying markets. Using data from a global survey of derivatives dealers and other sources, this article estimates the volume and potential impact of such hedging by U.S. dollar interest rate options dealers. In our analysis, we address two questions: First, are dealers' hedge adjustments large enough to affect trading volume and liquidity in the most common hedging instruments? Second, what effects might potential hedging difficulties have on risk premia in options prices and the structure of the market for over-the-counter interest rate options? In addressing these questions, we also consider whether dealers' dynamic hedging transactions have the potential to amplify price shocks.

We find that, on the whole, transaction volume in the underlying fixed-income markets is large enough to enable dealers to manage the risks incurred through their intermediation of price risk in the interest rate options market. Indeed, at shorter maturities, turnover volume in the most liquid hedging instruments is more than large enough to absorb the transaction volume generated by dealers' dynamic hedging. For medium-term maturities, however, an unusually large interest rate shock could cause the hedging of exposures in this segment of the yield curve to generate trading demand that is high relative to turnover volume in the more liquid trading instruments. Dealers then face a risk management trade-off between reducing price risk or incurring the liquidity costs of immediately rebalancing their hedge positions. However, only very large interest rate shocks, such as those occurring during a currency crisis or a period of high inflation, are likely to present dealers with this hedging problem.

In addition to analyzing hedging volume, we examine the term structure of options premia to assess whether option prices show any sign of potential hedging difficulties. We find an apparent risk premium in options prices at the medium-term segment of the yield curve that corresponds to the maturity range where our analysis of trading volume suggests that hedging difficulties might occur. This pattern in the term structure of options premia suggests that the liquidity risk in dynamic hedging may influence options pricing.

Table 1
OVER-THE-COUNTER INTEREST RATE OPTIONS DATA

NOTIONAL AMOUNTS REPORTED BY DEALERS, IN BILLIONS OF U.S. DOLLARS

	Bought Options			Sold Options		
Contracts with	U.S. Dollar Interest Rates	Other Interest Rates	Total	U.S. Dollar Interest Rates	Other Interest Rates	Total
Other dealers	529.4	726.5	1,255.9	576.1	681.9	1,258.1
Customers	431.6	340.6	772.2	690.4	398.1	1,088.4
Total	961.1	1,067.1	2,028.1	1,266.5	1,080.0	2,346.5

MARKET VALUES REPORTED BY DEALERS, IN BILLIONS OF U.S. DOLLARS

	Bought Options			Sold Options		
Contracts with	U.S. Dollar Interest Rates	Other Interest Rates	Total	U.S. Dollar Interest Rates	Other Interest Rates	Total
Other dealers	—	—	22.4	—	—	21.6
Customers	—	—	15.2	—	—	14.6
Total	20.8	16.7	37.6	19.4	16.8	36.2

MATURITY DISTRIBUTION OF U.S. DOLLAR INTEREST RATE OPTIONS, IN PERCENT

	Bought Options	Sold Options
Up to one year	30	29
More than one year and up to five years	58	56
More than five years	12	15

Source: Bank for International Settlements (1996).

Dynamic Hedging, Volatility of Financial Asset Prices, and Market Liquidity

An important question in any discussion of options hedging is whether the dynamic hedging of options in response to a price shock can introduce transactions large enough to amplify the initial price shock or to affect market liquidity. In asset price dynamics, such "positive feedback" occurs when an initial price change causes a shift in investor or trader demand that leads to a further change in price in the same direction. For example, a shift in investor sentiment in response to a sharp price decline can cause the sell-off of assets or widespread hedging of open positions—outcomes that would drive prices down further. The hedging of options also has the potential to cause positive feedback because dealers typically adjust their hedge positions by selling (buying) the underlying asset after its price falls (rises). These dynamic hedge adjustments in response to a fall in prices could introduce further downward pressure on prices.

Some observers cite the stock market crash of 1987—which occurred in the absence of any significant change in economic fundamentals—as an example of positive feedback dynamics. These observers suggest that the sharp fall in stock prices was intensified by portfolio insurance trading strategies that prescribe the sale (purchase) of stocks when prices fall (rise).[2] Although no empirical proof exists that positive feedback affects market prices, a number of papers (for example, Bank for International Settlements [1986], Grossman [1988], Gennotte and Leland [1990], and Pritsker [1997]) have suggested that dynamic hedging can cause positive feedback. In addition, Fernald, Keane, and Mosser (1994) discuss a possible example of positive feedback in the behavior of the term structure of interest rates.

If positive feedback is more than a theoretical possibility, then dynamic hedging would have the potential to amplify the volatility of asset prices when prices fall abruptly. Higher price volatility can in turn introduce other problems in financial markets. Most significantly, volatility can heighten uncertainty about credit risks and disrupt the intermediation of credit. For example, during the 1987 stock market crash, the increase of credit exposures in securities and margin settlement caused liquidity and funding problems for securities firms (see Bernanke [1990]). The potential for such financial market disruptions makes it worthwhile to consider the relationship between dynamic hedging and positive feedback in asset prices.

Dynamic hedging can also have implications for market liquidity. The financial innovations that have broadened the scope of financial intermediation to include the intermediation of price risks are positive developments that might be expected to lower risk premia in asset prices. Some of these forms of intermediation, however, rely on the ability of dealers to manage their risks dynamically. In the absence of market liquidity—which makes dynamic risk management possible—dealers would exact higher premia for their intermediation services. Some investors and fund managers may also rely on market liquidity in their investment and risk management strategies. If significant numbers of economic agents are relying on the liquidity of the core trading markets, either directly or indirectly, then part of the risk premia in financial asset prices might depend on assumptions about the robustness of that market liquidity. A sudden realization by investors and dealers that expectations of market liquidity were overly optimistic could lead to a sharp adjustment in asset prices. For this reason, assessments of the potential impact of dynamic hedging and risk management strategies on market liquidity are particularly useful. A related question is whether such dynamic risk management strategies by individual risk managers would be feasible in the aggregate during periods of extreme price volatility.[3]

An important question in any discussion of options hedging is whether the dynamic hedging of options in response to a price shock can introduce transactions large enough to amplify the initial price shock or to affect market liquidity.

A Brief Description of the Estimation

The analysis in this paper is based on data from the 1995 Central Bank Survey of Foreign Exchange and Derivatives Market Activity (Bank for International Settlements 1996), market growth data from the surveys of the International Swaps and Derivatives Association (ISDA), and historical interest rate data. The central bank survey reports the global market totals of outstanding derivatives contracts at the end of March 1995. The over-the-counter options data in the survey include notional amounts and market values of outstanding contracts, broken down by bought and sold options. A key part of our analysis is the derivation of strike prices that are consistent with the notional amount and market value data from the survey.[4]

Our estimation of dealers' hedging transactions has three principal steps. First, using the notional amount and rough maturity data from the central bank survey and market growth rates from the ISDA surveys, we estimate the distribution of notional amounts over maturities and origination dates. Next, we combine the estimated notional amounts at each origination date with historical interest rate data to estimate options strike prices that are consistent with the market values reported in the central bank survey and with historical interest rates. Finally, we use these strike prices to estimate the price sensitivity of a portfolio consisting of all dealers' interest rate options. Specifically, given the estimated strike prices, we calculate the delta of the global portfolio, that is, the change in the portfolio's value relative to changes in forward interest rates. The delta and its sensitivity to interest rate changes give us an estimate of dealers' hedge demands and dealers' hedge adjustments to interest rate shocks. (For a detailed description of the data and estimation, see the appendix.)

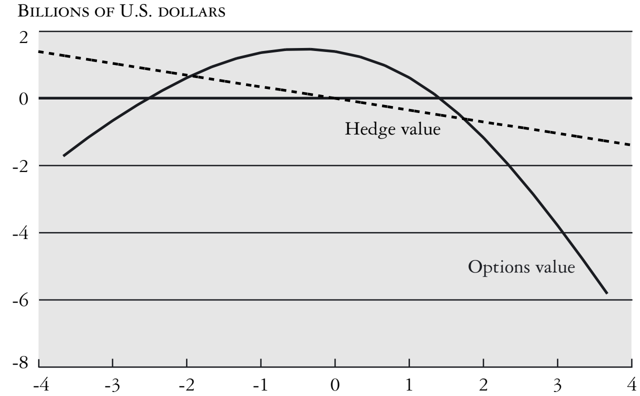
Estimated Price Risk in the Global Dealer Portfolio

We begin our analysis by using the estimated strike prices to derive the value of the global dealer portfolio of options at different interest rates (the solid line in Chart 1). The value of the options portfolio at the prevailing interest rates is the net market value reported in the central bank survey (Table 1, middle panel). The values at the indicated changes in interest rates are the option values calculated from the estimated strike prices. Chart 1 also shows, as a mirror image, the value of a hedge position that provides a delta-neutral hedge of the options at the initial interest rates (the dashed line). The hedge position is derived by using the estimated strike prices to calculate the price sensitivity (the delta) of the options portfolio. The estimated price sensitivity is used to construct a hedge position in fixed-income securities whose gain or loss in value offsets the change in value of the options portfolio for small changes in interest rates in either direction. The chart reveals a number of interesting facts about the dealers' portfolio of options.

First, at prevailing interest rates, the net value of the dealers' portfolio is positive. Although in notional amounts dealers sell more options than they purchase, at prevailing interest rates the bought options have higher market values than the sold options (Table 1, top and middle panels). This relationship between the notional amounts and the market values implies that the options sold to customers have a lower degree of moneyness than options purchased from customers (for definitions of terms, see box). The strike prices we estimate show the same relationship: relative to

Chart 1

Options and Hedge Values as a Function of Interest Rate Changes



Source: Author's caculations.

Notes: The hedge value is the mirror image of the value of a hedge position that provides the dealer with a delta-neutral position at the initial interest rate. The hedged portfolio has a positive value when the option value (the solid line) is above the hedge value (the dashed line).

swap interest rates at origination, sold options have estimated strikes that are out-of-the-money, while options purchased from customers are slightly in-the-money (see appendix). Because dealers are net sellers of options, however, large interest rate shocks will drive the sold options into-the-money, causing them to gain value, and as a result, the total value of the sold options will exceed the bought options' value. Hence, if the portfolio is not hedged, the aggregate dealers' portfolio value becomes negative when interest rates rise more than 125 basis points.

Options Terms and Concepts

Moneyness

An option's moneyness is a measure of its payoff at expiration. An option's payoff is defined relative to a specified level of the underlying asset's price called the *strike price*. For a call (put) option, the option is *in-the-money* at expiration when the asset price is above (below) the strike price, and the option pays the difference. When the asset price is below (above) the strike price at expiration, the call (put) option pays nothing, and the option is said to be *out-of-the-money*. An option's varying sensitivity to price risk is a result of the asymmetry in an option's payoff. An option is *at-the-money* when the underlying asset's price is equal to the strike price, and a call (put) option's moneyness is higher when the underlying asset's price is higher (lower).

Interest Rate Caps

Caps and floors are options on interest rates. In an interest rate cap (floor), if the interest rate at expiration of the contract is above (below) the strike rate specified in the contract, the buyer receives the difference, and nothing otherwise. Caps and floors are variations of call and put options. In terms of fixed-income securities, a cap is equivalent to a put option on a bond; in terms of interest rates, a cap is equivalent to a call option on interest rates. A cap (put option on a bond) gains value when interest rates rise (bond prices fall).

Varying Sensitivity to Price Risk and Positive Feedback

A call option's value increases by an amount smaller than the increase in the value of the underlying asset because there is always some probability that the price of the underlying asset will fall below the strike price at expiration, rendering the option worthless. As the underlying asset's price rises, this probability becomes smaller, and the value of the option becomes more sensitive to changes in the underlying asset's price. To compensate for this increase in the price sensitivity of a call option, a hedge position in the underlying asset must be made larger after the price of the underlying asset rises. This adjustment in the hedge position introduces the potential for positive feedback in price dynamics because the hedge adjustment is to buy (sell) the underlying asset after its price rises (falls).

Hedge Adjustments and Options Prices

As the value of the underlying asset rises, the writer of a call option must make the hedge position larger to ensure that its value is sufficient to cover the rising option exposure. As the value of the underlying asset falls, the hedge position must be reduced in size to ensure that the writer of the option is not left holding the underlying asset when the option expires out-of-the-money. Thus, the hedge adjustments in dynamic hedging involve buying the underlying asset after the price goes up and selling it after the price goes down. The cumulative cost of these "buy high, sell low" hedge adjustments equals the value of the option (for further discussion of option hedging, see Hull [1993]).

Volatility and Options Risk

The path-breaking option-pricing models developed more than two decades ago rely on continuous hedge adjustments to construct a dynamically hedged portfolio of underlying assets that perfectly replicates the payoff of an option (under the assumption that volatility remains constant). This ability to replicate the option means that the option does not contain unique risks, and, therefore, its value can be derived straightforwardly from the probability distribution of the underlying assets by using a risk-neutral expected value calculation. In practice, however, continuous hedge adjustments are not possible, and the difficulty in constructing a hedge portfolio that would perfectly replicate an option leaves the writer of an option with a unique and unhedgeable volatility risk.

Implied Volatility and Volatility Smiles

Market prices of options differ in characteristic ways from theoretical prices derived from benchmark pricing models, depending on the options' moneyness. These differences are manifested as differences in the implied volatility of the underlying asset when the benchmark model is used to infer the volatility of the underlying asset from the observed market price of the option. Options that are either deep out-of-the-money or deep in-the-money typically are priced in the market as if they had higher volatility in the log-normal distribution embedded in the benchmark pricing model. (This implied volatility pattern is called the volatility smile.) By incorporating these implied volatility differences in the benchmark pricing model, analysts can use the model to generate observed market prices.

Second, Chart 1 shows that a static hedge can only protect against small interest rate shocks. For a small change in interest rates, the change in the value of the options portfolio is offset by a corresponding change in the value of the hedge position. For a large interest rate change, however, the change in the value of the hedge position cannot offset the change in value of the options portfolio. Indeed, the hedged portfolio value turns negative where the options value and the hedge value intersect. Thus, with only a static hedge in place, the value of the hedged portfolio will turn negative after a large interest rate shock—specifically, an interest rate increase of more than 175 basis points. Dynamic adjustments to the hedge position as interest rates change, however, can prevent such an adverse outcome.

Third, to hedge against interest rate changes fully, dealers must adjust their hedge position after an interest rate shock. This adjustment compensates for the fact that the option portfolio's value falls at an increasing rate as interest rates rise.[5] As Chart 1 shows, without the hedge adjustment, the gain in value of the initial hedge position will no longer compensate for the decline in value of the option portfolio if interest rates continue to rise. This need to adjust the hedge position dynamically as interest rates change introduces the potential for positive feedback. Because the required hedge is a short position in fixed-income securities, the hedge adjustment to an increase in interest rates will introduce additional sales into the fixed-income market and may contribute further upward pressure on interest rates (by driving bond prices lower).

To hedge against interest rate changes fully, dealers must adjust their hedge position after an interest rate shock. This adjustment compensates for the fact that the option portfolio's value falls at an increasing rate as interest rates rise.

Finally, Chart 1 suggests that not all dealers can hedge their options exposures with offsetting exposures within their firms. The conventional view of financial institutions' interest rate risk profiles holds that these firms have a structural long position in the fixed-income market. That is, they have a firmwide exposure to rising rates. The negative slope of the options value curve at the prevailing forward rates, however, shows that the aggregate dealer portfolio of options has an exposure to rising interest rates as well. Thus, because the options portfolio and the other portfolios are exposed to rising rates, dealers as a group cannot hedge their net options exposures with offsetting structural positions in other parts of their firms. Although some dealers may rely on offsetting exposures elsewhere in their firms to hedge their options position, Chart 1 suggests that dealers as a group cannot hedge internally.

Estimated Scale of Dealers' Dynamic Hedging

A comparison of the size of dealers' hedge adjustments and transaction volume in the most common hedging instruments enables us to assess the market impact of dealers' hedging. As an option's moneyness increases after a price shock, the sensitivity of its value to further changes in prices increases. Thus, to maintain an option portfolio's exposure to price risk within a given limit, a dealer must adjust the hedge position after a price shock to allow for the change in the options' price sensitivity. For a given interest rate shock, we estimate the change in the hedge position required to restore the portfolio's price sensitivity (the delta) to its initial level. This hedge adjustment is the incremental demand of dealers for hedge instruments after an interest rate shock, if we assume that dealers maintain their exposure to price risk at some initial comfort level.

In our analysis of dealers' dynamic hedging, we make a number of assumptions. First, we assume that *customers* do not dynamically hedge their options positions because doing so would negate the investment or hedging objective that motivated the purchase of the option. Thus, we need consider only dealers' hedging demands. Second, we assume that interdealer options do not result in a net increase in dealers' hedge demands because they create only offsetting exposures among dealers.[6] Thus, we calculate dealers' net hedge requirements from dealers' contracts with customers. Finally, to calculate our benchmarks of dealers' hedging demands, we assume that dealers match the maturity of an option's interest rate exposure with the interest rate maturity of its hedge. For this reason, our benchmark hedge estimates are exact hedges that do not have yield curve or correlation risk. (For further discussion of these assumptions, see the appendix.)

The interest rate shocks we use in our estimates are increases in forward interest rates of 25 and 75 basis points. These interest rate changes are consistent with historical experience. For example, consider forward interest rates in the four-to-seven-year segment of the yield curve during the period 1991-95. For that period, a change of 25 basis points is slightly less than the largest daily change, and a change of 75 basis points is slightly less than the largest two-week change. During the last two decades, the ten largest daily changes in forward rates in the medium-term segment of the yield curve ranged from 60 to 100 basis points. At the short-term end of the yield curve, the ten largest daily changes in the three-month Treasury bill rate ranged from 80 to 130 basis points. These episodes of extreme volatility occurred between 1979 and 1981.

Estimates for the Most Common Hedging Instruments

In the U.S. dollar fixed-income market, options dealers executing their hedges can choose from a wide range of fixed-income instruments such as futures contracts, forward rate agreements (FRAs), interest rate swaps, interbank deposits, and Treasury and other bonds. These instruments, however, are imperfect substitutes because they have different credit risks, liquidity, and transaction costs. These differences create a need and an opportunity for intermediation. Dealers who provide risk management services to the markets take on and manage the risks and costs resulting from holding portfolios of such imperfect substitutes. When dealers have enough time to hedge a position or replace an initial hedge with a cheaper or better instrument, they can usually keep their exposure to price risk within manageable limits while still earning a profit from intermediation. When an immediate hedge adjustment in large volume is needed, however, dealers' hedging alternatives are more limited. For example, although the market for interest rate swaps is very large and becoming increasingly liquid, the daily turnover volume of swaps is still very small relative to outstanding contracts. The turnover of swaps is also small compared with turnover in the Eurodollar futures markets.[7] This difference in turnover volume suggests that swaps are more likely used to hedge structural or longer term exposures than to hedge positions that require frequent adjustment. Consequently, for dynamic hedge adjustments,

For dynamic hedge adjustments, dealers are likely to use the most liquid instruments as hedging vehicles. In the U.S. dollar fixed-income market, these instruments are Eurodollar futures, Treasury securities, *and* Treasury futures.

dealers are likely to use the most liquid instruments as hedging vehicles. In the U.S. dollar fixed-income market, these instruments are *Eurodollar futures, Treasury securities,* and *Treasury futures*.[8]

Hedging with Eurodollar Futures

Our estimate of dealers' hedging demands suggests that at shorter maturities the Eurodollar futures market is more than large enough to accommodate dealers' hedging—even when large interest rate shocks occur. For the hedging of longer maturity exposures, however, the Eurodollar futures market appears able to accommodate only the hedging of residual exposures, that is, marginal hedge adjustments and exposures that remain after the use of other hedging instruments.

As Tables 2 and 3 show, at maturities of up to one year, daily turnover volume exceeds the estimated hedge adjustment even when forward rates increase by as much as 75 basis points.[9] At longer maturities, however, the estimated hedge adjustments are sometimes larger than turnover volume. For a 25-basis-point change in forward rates, the largest daily turnover volume of Eurodollar futures contracts exceeds the estimated hedge adjustments for maturities out to five years and for maturities between eight and ten years (Table 2). For a 75-basis-point change in forward rates, however, the largest daily trading volume exceeds the estimated hedge adjustment out to only two years' maturity (Table 3). For small interest rate changes of, say, 10 basis points, daily turnover volume exceeds the estimated hedge adjustment at all maturities.

A comparison of the hedge estimates to another benchmark—the difference between the market's largest daily volume and its average volume—yields a similar conclusion. For maturities of up to two or three years, the surge in the largest daily volume exceeds the estimated hedge adjustment for a 25-basis-point change in forward rates (Table 2). For a 75-basis-point change in forward rates, however, the surge in volume exceeds the estimated hedge adjustment only out to maturities of a year and a half (Table 3). Thus, in response to a large interest rate shock, hedging volume at maturities beyond two years would be larger than daily turnover volume.

The stock of outstanding Eurodollar futures contracts also suggests that the market can support dealers' hedge adjustments. Our estimated hedge adjustments are smaller than the stock of outstanding futures contracts at all maturities. Even in the case of hedge adjustments in response to a 75-basis-point change in forward rates, the estimated hedge adjustment at most maturities is much less than half the outstanding futures contracts (Table 4). This result, along with our analysis of turnover volume, suggests that difficulties executing hedge adjustments are likely to be liquidity problems. That is, at the medium-term maturities, the Eurodollar futures market would have difficulty accommodating the entire hedging volume immediately, but the hedge adjustments could be absorbed over time.

So far, we have considered whether turnover volume is large enough to absorb transactions from *adjustments* to hedge positions in response to a price shock. Another consideration, however, is how large the hedge *position* is relative to outstanding contracts in the market. For the estimated

Table 2
CHANGE IN HEDGE POSITION FROM 25-BASIS-POINT INCREASE IN FORWARD RATES AND THE DAILY TURNOVER VOLUME OF EURODOLLAR FUTURES
Billions of U.S. Dollars

		Largest Daily Volume		Average Daily Volume		Difference between Largest and Average Daily Volume	
Maturity (Years)	Change in Hedge Position	First Contract	Second Contract	First Contract	Second Contract	First Contract	Second Contract
0.5	-6.3	**374.0**	**334.1**	**115.7**	**148.4**	**258.2**	**185.7**
1	-9.2	**260.9**	**135.2**	**92.1**	**35.8**	**168.8**	**99.4**
1.5	-7.7	**55.1**	**39.7**	**20.0**	**14.0**	**35.1**	**25.7**
2	-5.7	**26.9**	**18.9**	**9.4**	**6.0**	**17.5**	**13.0**
2.5	-4.6	**9.2**	**7.5**	4.0	3.3	**5.2**	4.3
3	-3.7	**7.3**	**4.5**	2.7	1.9	**4.6**	2.5
3.5	-3.1	**3.9**	2.6	1.5	1.3	2.4	1.3
4	-2.6	**2.7**	**3.3**	1.2	1.1	1.5	2.2
4.5	-2.1	**2.4**	**2.3**	0.9	0.8	1.5	1.5
5	-1.9	**2.0**	1.4	0.8	0.5	1.2	1.0
5.5	-1.6	1.3	**2.4**	0.2	0.2	1.1	**2.2**
6	-1.4	1.3	1.3	0.2	0.2	1.0	1.1
6.5	-1.2	1.0	1.2	0.2	0.2	0.9	1.1
7	-1.0	**3.3**	0.7	0.2	0.1	**3.1**	0.6
7.5	-0.9	0.6	**1.2**	0.1	0.1	0.5	**1.1**
8	-0.6	**0.8**	**3.7**	0.1	0.1	**0.7**	**3.5**
8.5	-0.4	**1.2**	**1.2**	0.1	0.1	**1.1**	**1.1**
9	-0.3	**1.2**	**1.7**	0.1	0.1	**1.1**	**1.6**
9.5	-0.1	**1.0**	**0.7**	0.1	0.1	**0.9**	**0.6**
10	—	1.2	1.0	0.1	0.0	1.1	1.0

Source: Author's calculations.

Notes: Hedge estimates are based on data as of the end of March 1995. Contract volume is for the first half of 1995. Bold type indicates that the contract volume exceeds the change in hedge position. Negative values indicate a short position. Because the futures contracts are contracts on a three-month interest rate, the hedge for each six-month exposure requires two back-to-back contracts ("first contract" and "second contract" in the table).

hedge position at longer maturities, the Eurodollar futures market is not large enough to accommodate all of the hedge demands that would be generated by a fully delta-neutral hedging strategy, particularly for exposures beyond four or five years (Table 4). Rather, at longer maturities, the Eurodollar futures market can accommodate only marginal hedge adjustments and the hedging of exposures that remain after the use of other hedging instruments.

Hedging with Treasury Securities and Treasury Futures

The Treasury securities market and the market for futures contracts on Treasuries—which are both large and highly liquid—are ideal complements to the Eurodollar futures market for dealers' hedging needs. To estimate the hedge of an exposure to forward rates between five and ten years' maturity, we assume that dealers' hedges consist of a short position (the sale of a borrowed security) in the ten-year note and a long position in the five-year note. This hedge can be executed with either cash market securities or futures contracts. The position is a hedge of the exposure between five and ten years' maturity because the long and short positions extinguish exposures to forward rates below five years, leaving only the exposure to forward rates beyond five years.

The Treasury cash and futures markets are generally large enough to accommodate dealers' dynamic hedging (Table 5). The estimated hedge adjustment is less than the combined daily turnover volume of on-the-run securities[10] and Treasury futures even for large interest rate shocks. However, dealers' hedging demand could be significant relative to the size of the markets. For example, the estimated hedge adjustment to a 75-basis-point shock could be as large as 21 percent of the combined average daily turnover in the Treasury futures and interdealer on-the-run cash markets, and almost 10 percent of the outstanding stocks of the on-the-run securities and futures contracts. Moreover, the estimated hedge position could be as large as a third of total outstanding contracts in the two markets. In sum, the Treasury cash and futures markets significantly expand the pool of fixed-income

Table 3

CHANGE IN HEDGE POSITION FROM 75-BASIS-POINT INCREASE IN FORWARD RATES AND THE DAILY TURNOVER VOLUME OF EURODOLLAR FUTURES

Billions of U.S. Dollars

		Largest Daily Volume		Average Daily Volume		Difference between Largest and Average Daily Volume	
Maturity (Years)	Change in Hedge Position	First Contract	Second Contract	First Contract	Second Contract	First Contract	Second Contract
0.5	-31.9	**374.0**	**334.1**	**115.7**	**148.4**	**258.2**	**185.7**
1	-31.2	**260.9**	**135.2**	**92.1**	**35.8**	**168.8**	**99.4**
1.5	-23.7	**55.1**	**39.7**	20.0	14.0	**35.1**	**25.7**
2	-17.2	**26.9**	**18.9**	9.4	6.0	**17.5**	13.0
2.5	-13.6	9.2	7.5	4.0	3.3	5.2	4.3
3	-11.0	7.3	4.5	2.7	1.9	4.6	2.5
3.5	-9.0	3.9	2.6	1.5	1.3	2.4	1.3
4	-7.6	2.7	3.3	1.2	1.1	1.5	2.2
4.5	-6.2	2.4	2.3	0.9	0.8	1.5	1.5
5	-5.5	2.0	1.4	0.8	0.5	1.2	1.0
5.5	-4.7	1.3	2.4	0.2	0.2	1.1	2.2
6	-4.1	1.3	1.3	0.2	0.2	1.0	1.1
6.5	-3.5	1.0	1.2	0.2	0.2	0.9	1.1
7	-3.0	**3.3**	0.7	0.2	0.1	**3.1**	0.6
7.5	-2.4	0.6	1.2	0.1	0.1	0.5	1.1
8	-1.9	0.8	**3.7**	0.1	0.1	0.7	**3.5**
8.5	-1.3	1.2	1.2	0.1	0.1	1.1	1.1
9	-0.7	**1.2**	**1.7**	0.1	0.1	**1.1**	**1.6**
9.5	-0.3	**1.0**	**0.7**	0.1	0.1	**0.9**	**0.6**
10	—	1.2	1.0	0.1	0.0	1.1	1.0

Source: Author's calculations.

Notes: Hedge estimates are based on data as of the end of March 1995. Contract volume is for the first half of 1995. Bold type indicates that the contract volume exceeds the change in hedge position. Negative values indicate a short position. Because the futures contracts are contracts on a three-month interest rate, the hedge for each six-month exposure requires two back-to-back contracts ("first contract" and "second contract" in the table).

instruments available to options dealers for their hedging needs. However, dealers' hedge demands could amount to a significant share of turnover volume and outstanding contracts in the Treasury on-the-run cash and futures markets.[11]

Alternative Hedging Estimates

The hedge estimates above do not account for changes in the volatility of interest rates. An option's hedge position against changes in the underlying asset's price, however, depends on the underlying asset's price volatility as well as on the asset's price level. Moreover, large changes in asset prices are often associated with higher implied volatilities in options. For this reason, alternative hedge adjustments were estimated assuming simultaneous volatility and interest rate level shocks (Table 6). Although the estimated hedge adjustment is larger, the difference does not appreciably change the conclusions because the difference from the base case is small relative to the turnover volume in the hedge instruments.

Table 4
DELTA-NEUTRAL HEDGE POSITION IN EURODOLLAR FUTURES CONTRACTS AND EURODOLLAR FUTURES CONTRACTS OUTSTANDING
Billions of U.S. Dollars

		Open Interest		
Maturity (Years)	Hedge Position	First Contract	Second Contract	Change in Hedge from a 75-Basis-Point Shock
0.5	38.3	**561.9**	**366.4**	-31.9
1	23.9	**279.7**	**222.0**	-31.2
1.5	2.8	**174.0**	**145.4**	-23.7
2	-4.0	**114.2**	**96.3**	-17.2
2.5	-9.8	**84.9**	**68.6**	-13.6
3	-13.4	**60.3**	**54.8**	-11.0
3.5	-16.4	**49.5**	**38.8**	-9.0
4	-17.9	**34.4**	**27.2**	-7.6
4.5	-20.2	**22.6**	14.5	-6.2
5	-18.9	12.9	9.5	-5.5
5.5	-18.8	7.5	7.7	-4.7
6	-18.4	6.2	5.9	-4.1
6.5	-17.5	6.7	6.8	-3.5
7	-15.1	6.8	4.5	-3.0
7.5	-12.6	3.8	2.5	-2.4
8	-9.6	1.6	2.2	-1.9
8.5	-6.2	1.8	1.8	-1.3
9	-3.4	1.7	2.0	-0.7
9.5	-1.4	0.8	0.9	-0.3
10	—	0.8	0.0	—

Source: Author's calculations.

Notes: The table reports delta-neutral hedge estimates and open interest as of the end of March 1995. Bold type indicates that the contract volume exceeds the change in hedge position. Negative values indicate a short position. Because the futures contracts are contracts on a three-month interest rate, the hedge for each six-month exposure requires two back-to-back contracts ("first contract" and "second contract" in the table).

DEALERS' HEDGE ADJUSTMENTS AND MARKET LIQUIDITY

Our estimate of dealers' hedging demands suggests that dealers might encounter hedging difficulties only for exposures beyond three or five years' maturity when large interest rate shocks occur. Together, the Eurodollar futures, on-the-run Treasury securities, and Treasury futures markets can absorb hedge adjustments to interest rate shocks as large as 25 basis points along the entire term structure. For example, for exposures between five and ten years' maturity, the estimated hedge adjustment to a 25-basis-point shock is only 7 percent of the combined turnover in the Treasury futures and interdealer on-the-run cash markets (Table 5).

For a large interest rate shock, however, such as a 75-basis-point shock to forward rates, dealers' dynamic hedge adjustments in the medium-term segment of the yield curve would generate significant demand relative to turnover in these hedging instruments. This demand would amount to 21 percent of the combined turnover in the Treasury futures and interdealer on-the-run cash markets (Table 5). In addition, the hedge adjustment in the three-

Table 5
HEDGE POSITION IN TREASURY SECURITIES AND FUTURES
Billions of U.S. Dollars

		Change in Hedge Position from an Interest Rate Shock of			On-the-Run Treasury Securities[a]		Treasury Futures[b]	
	Hedge Position	10 Basis Points	25 Basis Points	75 Basis Points	Outstanding	Daily Volume	Outstanding	Daily Volume
Five-year	13.0	0.4	1.0	2.9	13.2	9.0	19.7	5.1
Ten-year	-13.0	-0.4	-1.1	-3.3	13.8	6.0	25.8	9.2

Source: Author's calculations.

Notes: Hedge estimates are based on data as of the end of March 1995. Negative values indicate a short position.

[a]Outstanding amount as of the end of March 1995. Daily volume is estimated from interdealer trading volume (Fleming 1997).

[b]Five- and ten-year note contracts. Outstanding contracts are as of the end of March 1995. Daily volume is for the first half of 1995.

to-five-year maturity segment would be large relative to Eurodollar futures turnover volume (Table 3). If all dealers executed their hedge adjustments simultaneously, these transactions could have an impact on turnover volume and affect market liquidity. Moreover, in the presence of a large interest rate shock, other traders and investors might undertake transactions in the same direction as options dealers' hedge adjustments. All these demands together suggest that dealers wishing to adjust their hedge positions immediately could indeed encounter market liquidity problems.

Turnover volume in standard hedging instruments appears large enough to accommodate dealers' dynamic hedging in all but the most extreme periods of interest rate volatility.

Dealers can manage the impact on market liquidity by trading off price risk against the cost of immediacy or liquidity. For example, only part of the exposure opened up by a large interest rate shock might be hedged initially, with the remainder hedged over time. Dealers can spread the hedge adjustment over a number of days by executing a series of transactions that are small relative to daily turnover in the hedge instruments. This strategy reduces the market impact of the hedge adjustment but leaves the dealer exposed to some price risk until the hedging transactions are completed. Alternatively, by assuming some correlation risk, a dealer could also hedge the longer maturity exposures with the first three near-term futures contracts. The volume of these shortest maturity contracts is large enough to accommodate the hedging of longer maturity exposures easily, but returns on these contracts are less than perfectly correlated with longer maturity interest rates.

In another alternative, dealers could use an interest rate term structure model to design a hedge that avoids concentrated transactions at yield curve sectors with liquidity problems. For example, using a two-factor interest rate term structure model, a dealer could construct a hedge of exposures between five and ten years using a position in one-year bills and thirty-year bonds that replicates the exposure to the term structure factors that drive forward rates between five and ten years. Such hedges, however, are vulnerable to atypical price shocks not accounted for by the correlations in the term structure model.

Regardless of how the trade-off between price risk and the cost of immediacy or liquidity is executed, the terms of the trade-off depend on the volatility of interest rates. If volatility rises at the same time that liquidity is most impaired, then these hedging strategies could leave the firm exposed to higher than usual price risk.

These results suggest that transaction volume in the underlying fixed-income markets is large enough to enable dealers to manage the risks acquired from the intermediation of price risk in the interest rate options market. Turnover volume in standard hedging instruments appears large enough to accommodate dealers' dynamic hedging in all but the most extreme periods of interest rate volatility. For very large interest rate shocks, however, the hedging of exposures in the medium-term segment of the yield curve could lead to trading demand that is large relative to turnover volume in the more liquid trading instruments. Thus, for large interest

Table 6

CHANGE IN REQUIRED HEDGE POSITION FROM SIMULTANEOUS FORWARD AND VOLATILITY RATE SHOCKS

Billions of U.S. Dollars

Maturity (Years)	Interest Rate Shock	Volatility Shock	Interest Rate and Volatility Shocks
EURODOLLAR FUTURES			
0.5	-31.9	-6.0	-40.7
1	-31.2	-9.7	-38.7
1.5	-23.7	-8.7	-29.4
2	-17.2	-7.7	-22.6
2.5	-13.6	-6.2	-17.9
3	-11.0	-4.9	-14.4
3.5	-9.0	-3.9	-11.6
4	-7.6	-3.0	-9.5
4.5	-6.2	-2.2	-7.5
TREASURY SECURITIES AND FUTURES			
5	2.9	0.8	3.4
10	-3.3	-0.8	-3.8

Source: Author's calculations.

Notes: Hedge estimates are based on data as of the end of March 1995. The table assumes a 75-basis-point increase in forward rates. Volatility is assumed to increase by 25 percent relative to initial volatility levels at six months' maturity and by 8 percent at ten years' maturity. Negative values indicate a short position.

rate shocks, dealers' risk management decisions appear to be driven by a trade-off between price risk and the liquidity costs of immediate hedge rebalancing. Even so, for interest rate shocks typical of a low-inflation environment, the trade-off would need to be managed only for a short period of time.

THE EFFECT OF DYNAMIC HEDGING ON THE PRICES OF UNDERLYING ASSETS

Our results on the impact of dealers' dynamic hedging on prices in the fixed-income market are less clear. Any comprehensive assessment would need to account for the demands of other market participants as well. For example, investors whose demands are driven by macroeconomic fundamentals might undertake transactions in the opposite direction of dealers' dynamic hedging flows if interest rates were driven to unreasonable levels. If these investors constitute a sufficiently large part of the market, then their transactions could stabilize prices and reduce or even eliminate positive feedback dynamics (Pritsker 1997). These stabilizing investors, however, are not the only players. Traders who follow short-term market trends in "technical trading" strategies and speculators who anticipate the impact of positive feedback trading also participate in the market. These short-term traders could amplify the price impact of dealers' dynamic hedging because they would trade in the same direction as dealers' hedging transactions (see DeLong et al. [1990]). The ultimate impact of dealers' dynamic hedging would depend on the relative size of different types of market participants. For this reason, our analysis of the volume of dealers' hedging demands provides only a preliminary assessment of the potential for positive feedback because we have data on the hedging demands of dealers exclusively.

At shorter maturities, both the transaction volume and the outstanding stock of the most liquid trading instruments are much larger than dealers' dynamic hedging flows, so that the occurrence of positive feedback from dealers' dynamic hedging seems unlikely, even for very large interest rate shocks. At maturities beyond three years, however, if dealers fully rebalance their hedge positions, dynamic hedging in response to a large interest rate shock could be of significant volume relative to transaction volume and outstanding contracts in the most liquid trading instruments. At this segment of the yield curve, the potential for positive feedback when a very large interest rate shock occurs cannot be dismissed. The volume of dynamic hedging in response to an unusually large interest rate shock could be large enough to affect order flows and might temporarily affect the medium-term segment of the yield curve.

THE EFFECTS OF HEDGING DIFFICULTIES ON OPTIONS PRICING AND MARKET STRUCTURE

Our results suggest that the hedge adjustments of dealers in aggregate could be large relative to the size of the market for hedge instruments at the medium-term segment of the yield curve. Given this potential market impact, we consider whether market prices of interest rate options with medium-term maturities contain a premium to cover potential hedging difficulties. To look for evidence of a premium, we compare the implied volatility of interest rates derived from the market prices of options to historical interest rate volatility. An option's *implied volatility* is a volatility parameter in an options-pricing model that causes the model's option price to equal the option's actual price. If the market pricing of options contains a premium, we would expect the implied volatility to be large relative to other measures of the underlying asset's price volatility. Although by no means a comprehensive test, the simple comparison of the term structures of implied volatilities and historical volatilities provides a quick assessment of the possible existence of such a premium.

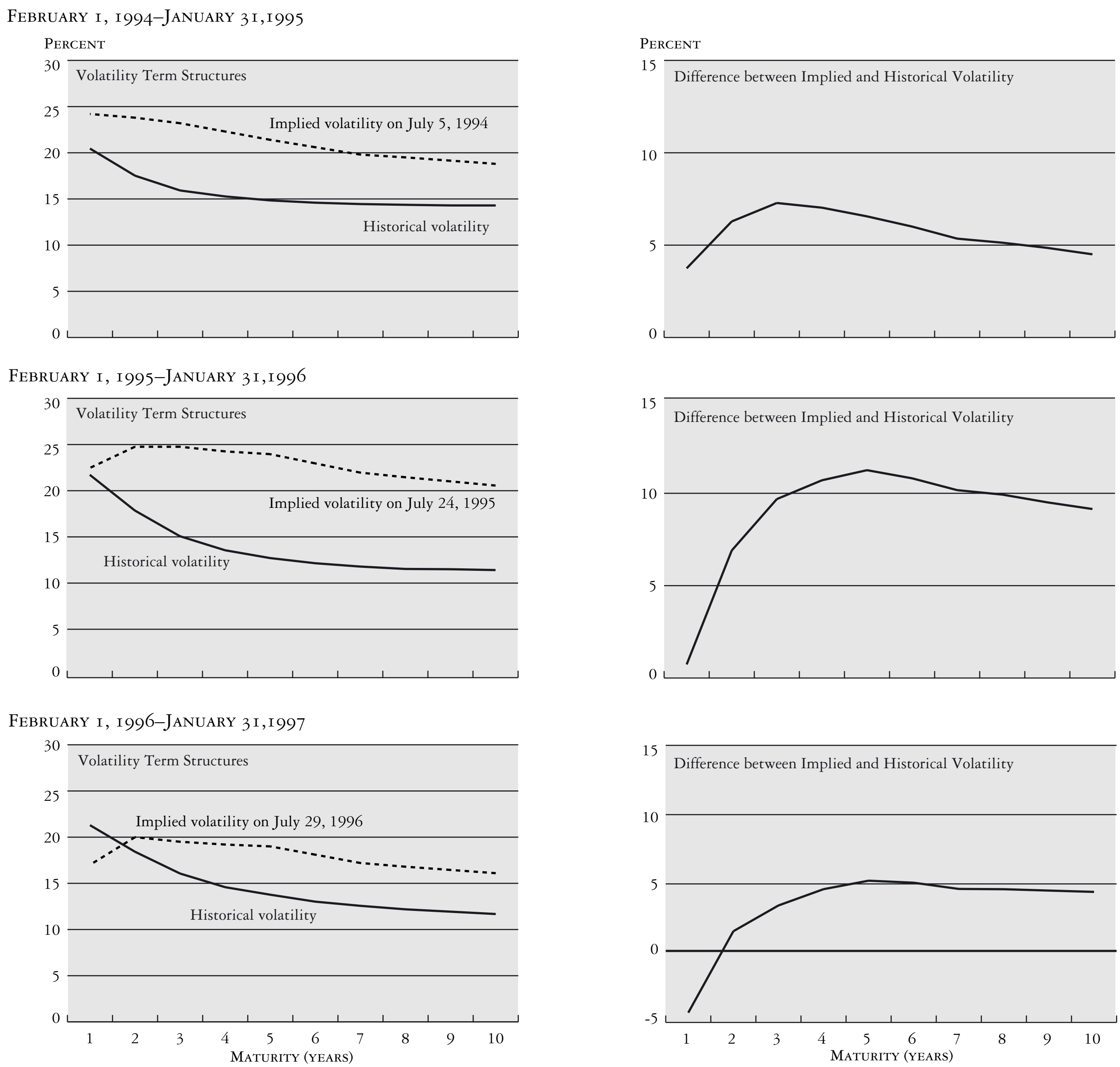
The difference between the term structure of implied volatility and the term structure of the historical volatility of forward Eurodollar interest rates is shown in Chart 2. Notably, the difference is greatest at the medium-term maturities (three to seven years), where the estimated hedge adjustments are large relative to the transaction volume of the hedge instruments. By contrast, the difference between implied volatility and historical volatility is smallest at short-term maturities (under two years). The estimated hedge adjustment relative to transaction volume in hedge instruments is also smallest at this maturity range. Although the difference between the historical and implied volatility term structures could reflect uncertainty about interest rate volatility in the medium-term segment of the yield curve, its shape is consistent with the existence of a premium for hedging difficulties. This apparent consistency between the term structure of options premia and our analysis of hedging volumes suggests a need for further research on how potential hedging difficulties may affect the term structure of interest rate options prices.

Volatility Risk and Hedging Costs

The change in the cost of hedging when interest rate volatility changes also affects the value of an option. Although the

Chart 2

Term Structures of Forward Interest Rate Volatility



Sources: Historical volatilities were derived using the yields on Eurodollar futures contracts as reported by DRI/McGraw-Hill. Implied volatilities are from *Derivatives Week*. In each panel, the historical volatility is for the period indicated.

exposure of an option's value to changes in the level of the underlying asset's price can be hedged, exposure to changes in the volatility of the underlying asset is not hedgeable with a linear fixed-income instrument such as a bond or a futures contract. Rather, an option's volatility risk can be hedged fully only with another option. Given that dealers as a group are net writers of options, their exposure to volatility risk is significant. If most customer options in the over-the-counter market are held to maturity, changes in volatility would

To look for evidence of a premium, we compare the implied volatility of interest rates derived from the market prices of options to historical interest rate volatility.

affect dealers through changes in the cost of hedging over the life of an option. Higher volatility would raise these hedging costs because it amplifies the costs of adjusting hedge ratios.[12] When volatility changes, the change in an option's value is equal to the expected change in hedging costs over the life of the option.

An estimate of the sensitivity to volatility shocks of the global dealer portfolio of interest rate options is shown in Chart 3. In the chart, the estimated strike prices are used to revalue the dealers' options portfolio for the indicated changes in volatility. An increase in volatility of approximately 35 to 40 percent causes the portfolio value to turn negative.[13] This change in value of the dealers' options portfolio is a measure of the volatility risk incurred by options dealers.

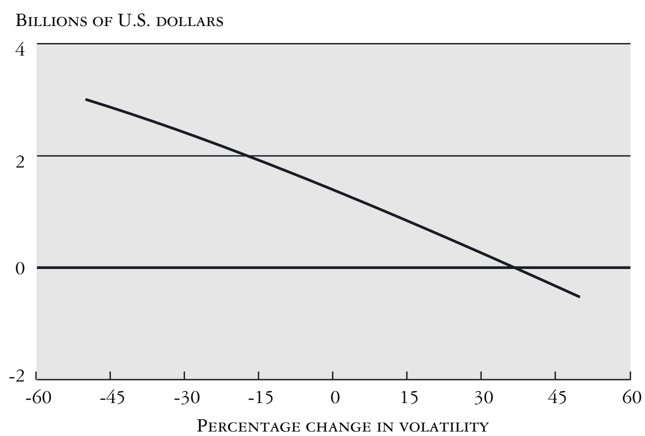
Hedging Difficulties, Volatility Risk, and Market Structure

Volatility risk and potential hedging difficulties may also affect the structure of the interest rate options market. In other derivatives markets, end-user needs are roughly balanced across buyers and sellers,[14] but in the over-the-counter interest rate options market, end-users are mostly buyers. As we noted earlier, dealers of U.S. dollar interest rate options have sold about 50 percent more options to customers than they have bought from customers. Thus, dealers are more willing than other investors to take on the volatility risk in selling an option. Given the wide range of financial assets and risks that investors are willing to acquire, why do they leave interest rate option exposure to dealers?

The concentration of interest rate option exposure among dealers implies that sellers of these options bear unique risks that are not present in the returns of underlying assets and that dealers are more willing to bear those risks. Volatility risk is one risk that is unique to options. Another is the difficulty in adjusting hedge positions as rapidly as required for the accurate hedging of an option's price risk.[15] The fact that dealers are more willing than other investors to sell interest rate options suggests that dealers are in a better position to bear the options' volatility and hedging risks. Dealers have two possible advantages in this area. First, they may be able to execute hedging transactions faster and at lower costs than other investors. Second, they may have other sources of income that offset the volatility risk in an option position. If dealers' income from market making in products other than options rises during periods of higher volatility, then that income will offset the increase in volatility risk from selling options. While some of that higher income would be compensation for the higher risk that dealers incur in making prices in volatile markets, any

Chart 3

Options Values as a Function of Volatility Changes



Source: Author's calculations.

remaining excess returns would offset the higher volatility risk from the options book. Such offsetting risks may explain why large interest rate options dealers also are market makers in a broad array of fixed-income products, and, for that reason, are willing to bear volatility risk at a smaller premium than other investors.

The evidence that market making in options *and* other products provides dealers with offsetting exposures to changes in volatility is not strong, however. For instance, even though the turnover volume of derivatives grew rapidly during 1994, dealers' trading income suffered from the bond market turbulence that occurred in that year. It has been reported that a significant part of the 1994 earnings decline occurred in dealers' bond and proprietary trading positions and not in their market-making activity.[16] Nevertheless, we lack detailed data on market-making income that would enable us to resolve with any certainty the question of offsetting exposures to volatility.

CONCLUSION

Our analysis suggests that transaction volume in underlying markets is large enough for dealers to manage the price and liquidity risks they incur through the intermediation of price risk in selling interest rate options. With the possible exception of the medium-term segment of the term structure, turnover volume in the most liquid hedging instruments is large enough to absorb dealers' dynamic hedging.

In the case of an unusually large interest rate shock at the medium-term segment of the term structure, the full rebalancing of hedge positions would generate hedging transactions that would be large relative to daily transaction volume in the most liquid medium-term instruments. In this case, dealers' risk management decisions would appear to be driven by a trade-off between price risk and the liquidity costs of immediate hedge rebalancing. For interest rate shocks of the size experienced in the last five years, dealers' hedge adjustments would be a small proportion of only a few days' worth of turnover volume, and dealers would need to manage the trade-off between liquidity and price risks only for a short period of time. For large interest rate shocks, however, such as those experienced by a country in the midst of a currency crisis or a period of high inflation, the hedging of exposures in the medium-term segment of the yield curve could lead to trading demand that is large relative to turnover volume in the more liquid trading instruments.

The ratios of estimated hedge adjustments to transaction volume in trading instruments at different maturities are consistent with the pattern we find in the term structure of option premia. The term structure of implied volatility shows an apparent risk premium for options at the medium-term segment of the yield curve, a segment that corresponds to the maturity range where hedging difficulties might occur. The structure of the over-the-counter interest rate options market is also consistent with the hypothesis that such hedging problems may exist. Despite investors' willingness to hold a wide variety of financial assets and risks, they choose to leave interest rate options exposures in the hands of dealers. This preference suggests that interest rate options sellers are exposed to risks that are not present in the returns of underlying assets. These risks are likely volatility and hedging-related risks, which may be managed more effectively by dealers than by other market participants.

The results presented in this article provide a preliminary assessment of the impact of dynamic hedging on market liquidity and price dynamics in the fixed-income market. As the appendix makes clear, limitations of the data make further investigation worthwhile. In addition, an estimate of the market excess demand function and the relationship between prices and quantities would be useful. Such an analysis, however, would require data that do not currently exist on investors' demands in addition to dealers' hedging demands. Nonetheless, comparing potential hedging demand with transaction volume in typical hedging instruments is useful in assessing the likelihood of positive feedback.

Appendix: The Estimation

The Data

Our primary source of data is the 1995 Central Bank Survey of Foreign Exchange and Derivatives Market Activity (Bank for International Settlements 1996). This survey of derivatives dealers worldwide reports global market totals of outstanding contracts as of the end of March 1995. The over-the-counter options market is a dealer market where all options contracts involve a dealer on at least one side of the contract. An option contract can either be a transaction between two dealers or a contract between a dealer and a customer. The central bank survey captured the entire over-the-counter options market by collecting data from the dealers that executed all contracts.[17] The options data in the survey include notional amounts, market values, and maturity data, broken down by bought and sold options, as shown in Table 1.[18] The options were also broken down by the survey's three counterparty categories: interdealer options, options bought from customers by dealers, and options sold to customers by dealers. Because reporters in the survey were derivatives dealers, interdealer transactions appear as both bought and sold options. In other words, an option bought by one dealer from another was reported as a bought option by one dealer and as a sold option by the other.[19]

Option Valuation

All options in the estimation are caps and floors on six-month interest rates. In accordance with the data from the International Swaps and Derivatives Association (ISDA), 73 percent of the options in the estimation are assumed to be caps, and the remainder are assumed to be floors. Although a small proportion of interest rate options are swaptions (19 percent at year-end 1994 in the ISDA data), for simplicity, we treat all options as either caps or floors.[20] Option values are calculated using Black's forward contract option model, the benchmark model used for implied volatility quotes for interest rate options (see Hull [1993]). The valuation uses the term structure of forward rates and the term structure of implied volatilities coinciding with the central bank survey data (end of March 1995).[21] To test the valuations, we also calculate the option values using different volatility structures. The baseline case assumes that caps and floors have identical implied volatilities that do not vary with moneyness. Alternative valuations using a volatility smile for options with different degrees of moneyness and higher implied volatilities for floors relative to caps did not affect our conclusions (see table below).

Maturity Distribution

To estimate the distribution of notional amounts over the remaining maturity and origination dates, we fit a quadratic function defined over original maturities to the remaining maturity data from the central bank survey and the ISDA

Volatility Assumptions: Change in Hedge Position from 75-Basis-Point Increase in Forward Rates
Billions of U.S. Dollars

	Assumption			
Maturity (Years)	Base	Cap/Floor Volatility	Volatility Smile	Cap/Floor Volatility and Smile
Eurodollar Futures				
0.5	-31.9	-31.5	-27.2	-26.8
1	-31.2	-31.2	-27.8	-27.7
1.5	-23.7	-23.7	-21.1	-21.0
2	-17.2	-17.2	-14.6	-14.5
2.5	-13.6	-13.6	-11.4	-11.3
3	-11.0	-10.9	-9.1	-9.0
3.5	-9.0	-9.0	-7.4	-7.4
4	-7.6	-7.5	-6.2	-6.2
4.5	-6.2	-6.2	-5.3	-5.3
Treasury Securities and Futures				
5	2.9	2.9	2.6	2.6
10	-3.3	-3.3	-2.9	-2.9

Source: Author's calculations.

Notes: Hedge estimates are based on data at the end of March 1995. Negative values indicate an increase in a short position. The base is the estimate using the assumptions in the text. For the cap/floor volatility assumption, option values were calculated with higher volatility for floors using cap and floor implied volatility differences reported by DRI/McGraw-Hill. For the volatility smile assumption, option values were calculated using a volatility smile derived from Eurodollar futures options prices.

market growth data. The three maturity categories in the central bank survey provide the three equations required to estimate the three parameters of the quadratic function. In the estimation, the market growth rates between origination dates are applied as a scaling factor to the quadratic function. We estimate separate maturity distributions for options purchased from customers and options sold to customers.[22]

In the estimated distribution, most outstanding contracts have less than four years' remaining maturity and have origination dates that fall within three years of the central bank survey date. The estimated distribution has a trough along the diagonal for caps with maturities at origination of between five and ten years. This feature of the distribution suggests that long maturity caps are originated at discrete maturities, specifically at the ten-year maturity.[23]

STRIKE PRICES

Strike prices are derived from historical yield curves and assigned to the options using the estimated distribution of notional amounts over origination dates. Because separate market values are not available for caps and floors, the estimation requires that a relationship between the strikes of caps and floors be imposed. The relationship assumed is that buyers (or sellers) of caps and floors have similar preferences regarding their options' moneyness. Under this assumption, if buyers of caps desire out-of-the money options because of their cheaper premia, then buyers of floors will also.

We implement the assumed relationship regarding the moneyness of caps and floors in three different ways. In all three approaches, the historical swap term structure at an option's origination date is our starting point. The first method is a proportional displacement of the strike rates from the historical swap term structure (in the same proportion, but opposite directions, for caps and floors). The other two methods are displacements of the strikes from the historical swap term structure with a constraint that caps and floors (of the same maturity and origination date) have equal premia (the second method) or equal deltas (the third method). Under the last two methods, the strikes for caps and floors can have different displacements from the swap term structure.

In these specifications, a cap will be out-of-the-money at origination when a floor is out-of-the-money. In each specification, the restrictions are applied to bought and sold options separately. The figures in the text are derived using the first approach, but similar results followed from the other specifications.[24]

ESTIMATED STRIKE PRICES AND OPTION VALUES

Given the strike price specification relative to historical yield curves, option values are calculated as functions of the displacement of the strike prices from the historical yield curves. The estimation then involves finding the displacement that produces option values equal to the market values observed in the central bank survey.

The objective of the estimation is to find values of the strike price displacement variables (A_b, A_s) such that

$$V_b(A_b) + v_D = v_b$$
$$V_s(A_s) + v_D = v_s,$$

where v_b and v_s are the observed market values of U.S. dollar options bought and sold by dealers (including interdealer options), and v_D is the market value of interdealer options. The functions $V(A)$ are the option values as functions of the displacement (A) of the strike prices from the historical term structures, and the subscripts indicate options bought (b) and sold (s) from customers.[25] In the proportional displacement specification of the strike price, the term A is a single variable. In the other two cases, the term A is a vector with two elements—the displacement for caps and the displacement for floors. In both cases, the additional equation required to solve for the two displacement variables is the equal premia or equal delta restriction in the specification of the strike prices.

In the option value equations above, the market value of U.S. dollar interdealer options is not available directly from the central bank survey data because market values are reported in aggregate—across counterparty types and across currencies (Table 1, middle panel). To find the displacement variables for bought and sold options (A_b and A_s) in the option value equations above, we first estimate the value of U.S. dollar interdealer options.

INTERDEALER OPTIONS

Estimates of the market value of U.S. dollar interdealer options are calculated in four different ways. In the first three methods we make assumptions about the strike rates of interdealer options: (1) interdealer options have strikes equal to swap rates (at-the-money strikes relative to the swap term structure); (2) interdealer options have the same strikes as options bought from customers; and (3) interdealer options have the same strikes as options sold to customers.[26] In the fourth method, we estimate the value of U.S. dollar interdealer options using the data reported in Table 1. In this method, the market value of interdealer options is distributed between U.S. dollar options and options on other currencies so as to minimize the discrepancy between the ratio of market value to notional amount for each currency and counterparty type and the ratio of the market value to the notional amount of the margin totals in the top and middle panels of Table 1.

The first and last alternatives produce comparable values for U.S. dollar interdealer options, while the other two do not. The estimation using the at-the-money assumption produces a value of interdealer options of $11.3 billion, while the last method results in a value of interdealer options of $10.9 billion. Strike prices that produce a value of $10.9 billion would be very slightly out-of-the-money at origination. The comparability of the estimates in methods one and four implies that interdealer options have strikes closer to at-the-money than do customer options. This result is plausible because dealers using the interdealer market to hedge their short volatility and negative gamma position would obtain more hedging benefit from at-the-money options. Such options have larger gamma and provide the most hedging benefit relative to their premia. The results reported in the text are derived using the fourth method. Despite the different estimates of interdealer option values, similar hedge estimates follow from all four alternatives.

CUSTOMER OPTIONS

For options sold to customers, estimated strikes consistent with observed market values are deep out-of-the money (relative to swap rates of comparable maturity) at origination. This result is plausible—customers buying options to hedge can acquire inexpensive protection against large interest rate shocks with deep out-of-the money options. For caps sold to customers, the estimated strike rates are 18 percent higher than swap rates of the same maturity at origination. The figure of 18 percent is comparable to the standard deviation of annual changes in interest rates, or two standard deviations of quarterly interest rate changes (six-month LIBOR rates from January 1991 to December 1995).

For options bought from customers, strike prices consistent with the observed market values are slightly in-the-money (relative to swap rates of comparable maturity) at origination. This relationship is the opposite of the relationship found for options sold to customers. Although this result might appear counterintuitive and could point to a problem in the estimation, it is consistent with market commentary in the early 1990s. Customers looking for "yield enhancement" during the low-interest-rate regime of the early 1990s acquired higher premia by selling interest rate caps with a higher degree of moneyness. While this higher yield is the market price or compensation for the expected payout of the option, investors speculating on the path of interest rates by selling options would obtain higher investment returns (or losses) per option by selling in-the-money options. In addition, investors who believed that the forward curve was an overestimate of the future path of spot rates would have sold options that were in-the-money relative to swap rates.

HEDGING ASSUMPTIONS

The final step in the estimation of dealers' hedge adjustments is the calculation of the delta and the change in delta of the global dealers' portfolio using the estimated distribution of notional amounts and the estimated strike prices. The analysis of dealers' hedging behavior relies on the following assumptions:

1. After an interest rate shock, dealers restore the net delta of their position to its initial level. Dealers may or may not fully hedge the initial delta of the options book, and whatever hedging is done initially may be accomplished either internally, with offsetting positions in the firm, or externally, with hedging transactions. These initial offsetting positions, either internal or external, are assumed to have a small gamma, so that changes in interest rates—and thus the options' delta—make additional hedging transactions necessary to return the portfolio's net delta to its original level.

2. An option exposure to a period t interest rate is hedged with an instrument that also has exposure to the period t interest rate—there is no basis risk in hedged positions. Using this assumption, we calculate a separate hedge ratio for each maturity's exposure.

3. Customers do not hedge their options positions. Customers who have bought or sold options are assumed not to hedge, because doing so would negate whatever investment or hedging objective the options were used for. Customers who have sold options to dealers presumably did so for speculative "yield enhancement" or intertemporal income shifting. The costs of delta-hedging the options would negate that investment objective. Customers who have bought options from dealers for hedging purposes would not hedge the option because doing so would expose the underlying position that the option was hedging. Thus, the impact of dynamic hedging is assessed using the aggregate dealers' position.

   If customers were to hedge their options, perhaps as a result of a reassessment of risks, then the market impact of dealers' hedge adjustments would be smaller because these adjustments would be offset by customer hedges. Because most of our results support the claim that the market impact of dealers' hedging is small relative to the size of the market, dropping the assumption would strengthen our conclusion that the markets for typical hedging instruments are sufficiently large for dealers to manage the price risk acquired from market making in options.

4. Interdealer options have no effect on dealers' net demand for hedge instruments. Using this assumption, interdealer options can be ignored, and the net hedge position and hedge adjustment of dealers in the aggregate can be calculated from dealers' contracts with customers. This assumption is reasonable when interdealer options are executed to reallocate customer exposures among dealers or to take a position in volatility risk but not directional interest rate risk. In the first case, the interdealer option that passes a customer exposure from one dealer to another does not create additional net option exposure for dealers in the aggregate. Thus, dealers' net hedge demands would be unaffected by such interdealer options.

   The second type of interdealer trading that is consistent with this assumption is position taking on changes in interest rate volatility. This trading strategy entails the hedging of directional interest rate risk. If executed by dealers on the two sides of an interdealer trade, such hedges would offset each other in the market, with no impact on the net dealer hedge amount.

An important justification for the presumption that dealers' position taking in options is a position on volatility changes is the fact that dealers wishing to take directional exposures to interest rate risk could do so less expensively with instruments other than options.

ROBUSTNESS CHECKS

To test our results, we derive estimates of dealers' hedging using alternative assumptions about implied volatility, the structure of strike prices, and other restrictions. The variation in hedge demands across these assumptions is small relative to turnover volume in the hedge instruments and does not change our conclusions. The results under these alternative assumptions are available in Kambhu (1997, Tables 5-8).

Although the results are robust to alternative assumptions, they might be influenced by certain features of the central bank survey data. First, dealers might have had options positions that were not reported in the central bank survey. Index amortizing interest rate (IAR) swaps, for example, might have been reported as swaps instead of options. These instruments were popular in the early 1990s, when investors were searching for yield enhancement. The extra yield in this instrument is the premium for a written option embedded in the instrument's payoff structure. Most of the volume of IAR swaps, however, was in contracts of three years' or shorter maturity. By the time of the survey, outstanding volume was likely to have been too small to affect the results.[27]

In addition, the timing of the central bank survey may have caused the survey data to capture patterns in option strike rates, the mix of bought and sold options, or maturity that were unique to 1995. The survey in 1995 followed a period of low interest rates in the early 1990s and a shift to tighter monetary policy in 1994. Data from the ISDA surveys show that the over-the-counter interest rate options market grew rapidly in 1993 and 1995. Growth, however, was lower than usual in 1994. The interest rate swaps market, by contrast, grew rapidly in 1994, especially during the first half of the year. Whether these patterns affected the survey results can best be determined by replicating the study at some future date.[28]

ENDNOTES

The author is grateful for the helpful comments and suggestions of Young Ho Eom, Frank Keane, James Mahoney, and participants in workshops at the Bank for International Settlements and the Federal Reserve Bank of New York.

1. For additional discussion of the intermediation of price risk, see Kambhu, Keane, and Benadon (1996).

2. Gennotte and Leland (1990) summarize the debate surrounding the role of portfolio insurance in the 1987 stock market crash.

3. For further discussion of market liquidity and risk management, see Bank for International Settlements (1995, Chap. 2).

4. The ISDA data consist of notional amounts but not market values. As a result, the analysis in this article was not possible until the 1995 central bank survey supplied market value data as well as a breakdown of gross options positions by bought and sold transactions.

5. The value of an interest rate cap becomes more sensitive to changes in interest rates as rates rise (see box). In Chart 1, the dealers' portfolio value falls at an increasing rate because dealers are net sellers of options and thus incur increasing option liability as rates rise.

6. This assumption is reasonable when interdealer options are executed to reallocate customer exposures among dealers, or to take a position in volatility risk but not directional interest rate risk. For further explanation, see the hedging assumptions section in the appendix.

7. Turnover volume for U.S. dollar interest rate swaps at the time of the survey was $17 billion per day (Bank for International Settlements 1996). In contrast, Eurodollar futures turnover volume was $463 billion per day, and turnover of the five- and ten-year Treasury on-the-run securities and futures was $29 billion per day.

8. Exchange-traded options on futures contracts are also a potential hedging instrument. The survey data, however, show that dealers as a group have bought and sold roughly equal amounts of exchange-traded options. Thus, these instruments cannot be providing a net hedge to the aggregate dealer position, and dealers as a group must be relying on other hedging instruments.

9. In Tables 2, 3, and 4, the hedge for each six-month exposure requires two back-to-back futures contracts because the contracts are on a three-month interest rate. For example, in Table 2, in response to a 25-basis-point rise in interest rates, at the two-and-a-half-year maturity the hedge adjustment comprises a sale of $4.6 billion in each of the two back-to-back contracts that span the interval between two-and-a-half and three years. This amount is less than the turnover volume of $9.2 billion and $7.5 billion in the two contracts that match the maturity of the hedge position.

10. On-the-run securities, or the most recently issued securities, are the most liquid Treasury issues. As a security ages, a larger proportion of the issue tends to be held in long-term investment portfolios and thus is traded less frequently.

11. The cash market for the on-the-run Treasury security by itself appears too small to accommodate dealers' hedge demands. If dealers fully hedged their exposures beyond five years using five- and ten-year on-the-run issues, the required hedge position would be approximately equal to the outstanding amount of the on-the-run five- and ten-year notes (Table 5). The Treasury securities market, however, can still accommodate a significant share of this hedging demand in two ways. First, the lending of Treasury securities in the repo market allows a fixed stock of on-the-run Treasury securities to meet trading demands that exceed the size of the on-the-run issue. Through the repo market, a trader who sells a borrowed security to establish a short position enables another trader to establish a long position in the security. As a result, market participants' long positions in the security can be significantly larger than the outstanding stock of the security. Second, off-the-run issues can be used as long as they are available. Fleming (1997) reports that off-the-run securities account for approximately 24 percent of daily turnover in the interdealer market.

12. Dynamic hedging requires a dealer to buy the underlying asset after its price rises and to sell it after the price falls. The cost of implementing this "buy high, sell low" trading strategy is higher when price changes are more volatile.

13. In the five-year period beginning in 1991, the three largest changes in implied volatility for one-year options on Eurodollar futures were between 33 percent and 38 percent for two-week changes in implied volatility.

14. See Kambhu, Keane, and Benadon (1996).

15. For a study of how market prices of options are influenced by volatility and hedging risks, see, for example, Jameson and Wilhelm (1992) for the pricing of exchange-traded stock options.

16. See *Risk* (1994) and *Swaps Monitor* (1996).

17. The interest-rate-related options were predominantly caps and floors. The central bank survey also included data for over-the-counter options on traded interest rate securities (bond options). These options were not included in our analysis, because they amounted to less than 8 percent of options related to interest rates. Moreover, the bought and sold amounts

Note 17 continued
of these options were in rough balance, leaving dealers with very little residual hedging needs from positions in these options.

18. The notional amount of a derivative contract is a reference amount used to calculate the size of the cash flows between the counterparties to the contract. (These cash flows are determined by the price of an underlying asset.) The market value of a derivative contract is the net value of the cash flows to be exchanged between the counterparties over the remaining life of the contract. Notional amounts are a measure of contract size that is independent of the price of the underlying asset, while market values are a measure of contract size that is based on the market value of the transaction. Market values of derivative contracts are almost always a small proportion of the notional amount.

19. Because of reporting error, the bought and sold amounts of interdealer options reported in the survey are slightly different. To account for this reporting error, we average the bought and sold figures to arrive at the interdealer volume used in the estimation. This averaging should reduce the effects of the error.

20. The exclusion of swaptions is not likely to alter the article's conclusions for the following reasons. If a one-year option on a five-year swap was reported as a one-year option, then the swaptions would appear as shorter maturity options in the data. Hence, the true exposures of shorter maturity would be less than estimated, with the result that hedging demand for shorter maturity instruments would be smaller than estimated. This effect would only strengthen the conclusion that shorter maturity hedging volumes are small relative to transaction volume in Eurodollar futures. The swaptions, however, would add to the estimated hedging demand at longer maturities. Nevertheless, because swaptions make up only 19 percent of the market, the net increment to estimated hedging demand would not significantly change our conclusions. Rather, the effect would be to strengthen the conclusion that longer maturity hedging demand could be significant relative to transaction flows in longer maturity hedge instruments, but not so much larger as to overwhelm the market.

21. The implied volatility and forward interest rate data are from *Derivatives Week* (1995). The *Derivatives Week* data on forward rates and implied volatility are consistent with those implied by Eurodollar futures prices and Eurodollar futures options prices.

22. For further details, see Kambhu (1997).

23. To test whether the clustering at the ten-year maturity was a product of the quadratic function in the estimation, we derived an alternative estimate using a linear maturity distribution out to seven years and a separately estimated ten-year share. This alternative produced similar results for both the maturity distribution and the hedging volumes. In a further test, a linear distribution out to nine years with a separate ten-year share produced nonsensical results with negative values at the longer maturities in the linear segment of the distribution.

We also derived results with alternative maturity estimates. These alternatives did not change our conclusions (see Kambhu [1997, Table 7] for further details). The heavy distribution of notional amounts in the one-year remaining maturity range, which constrains the effects of the alternative estimation methods, may explain the robustness of the results.

24. Alternative strike price structures did not produce much variation in the hedge estimates relative to turnover volume in the hedge instruments and thus did not affect the conclusions. See Kambhu (1997, Table 5) for further details.

25. These functions are defined by the strike price specification and the estimated distribution of notional amounts. See Kambhu (1997) for further details.

26. In the first three methods, the estimation of interdealer market values relies on the assumption that the maturity structure of interdealer options is equal to the average of the bought and sold options' maturity distributions.

27. Cumulative volume of IAR swaps originating between 1990 and 1994 was about $100 billion to $150 billion in notional principal (Galaif 1993-94).

28. Beginning in June 1998, global derivatives market data similar to the 1995 survey will be collected on a semiannual basis by the Group of Ten central banks and published by the Bank for International Settlements.

REFERENCES

Bank for International Settlements. 1986. RECENT INNOVATIONS IN INTERNATIONAL BANKING. April.

———. 1995. ISSUES OF MEASUREMENT RELATED TO MARKET SIZE AND MACROPRUDENTIAL RISKS IN DERIVATIVES MARKETS. February.

———. 1996. CENTRAL BANK SURVEY OF FOREIGN EXCHANGE AND DERIVATIVES MARKET ACTIVITY. May.

Bernanke, Ben. 1990. "Clearing and Settlement during the Crash." REVIEW OF FINANCIAL STUDIES 3, no. 1: 133-51.

Board of Governors of the Federal Reserve System. 1996. FEDERAL RESERVE BULLETIN 82, no. 6 (July).

DeLong, J. Bradford, Andrei Shleifer, Lawrence Summers, and Robert Waldmann. 1990. "Positive Feedback Investment Strategies and Destabilizing Rational Speculation." JOURNAL OF FINANCE 45, no. 2 (June): 379-95.

DERIVATIVES WEEK. 1995. "Interest Rate Monitor." Vol. 4, no. 13 (April 3): 9.

Fernald, Julia, Frank Keane, and Patricia Mosser. 1994. "Mortgage Security Hedging and the Yield Curve." Federal Reserve Bank of New York QUARTERLY REVIEW 19, no. 2 (summer-fall): 92-100.

Fleming, Michael. 1997. "The Round-the-Clock Market for U.S. Treasury Securities." Federal Reserve Bank of New York ECONOMIC POLICY REVIEW 3, no. 2 (July): 9-32.

Galaif, Lisa. 1993-94. "Index Amortizing Rate Swaps." Federal Reserve Bank of New York QUARTERLY REVIEW 18, no. 4 (winter): 63-70.

Gennotte, Gerard, and Hayne Leland. 1990. "Market Liquidity, Hedging, and Crashes." AMERICAN ECONOMIC REVIEW 80, no. 5: 999-1021.

Grossman, Sanford. 1988. "An Analysis of the Implications for Stock and Futures Price Volatility of Program Hedging and Dynamic Hedging Strategies." JOURNAL OF BUSINESS 61 (July): 275-98.

Hull, John. 1993. OPTIONS, FUTURES, AND OTHER DERIVATIVE SECURITIES. 2d ed. Englewood Cliffs, N.J.: Prentice Hall.

Jameson, Mel, and William Wilhelm. 1992. "Market Making in the Options Markets and the Costs of Discrete Hedge Rebalancing." JOURNAL OF FINANCE 47, no. 2 (June): 765-79.

Kambhu, John. 1997. "The Size of Hedge Adjustments of Derivatives Dealers' U.S. Dollar Interest Rate Options." In ON THE MEASUREMENT OF AGGREGATE MARKET RISK. Basle: Bank for International Settlements, November.

Kambhu, John, Frank Keane, and Catherine Benadon. 1996. "Price Risk Intermediation in the Over-the-Counter Derivatives Markets: Interpretation of a Global Survey." Federal Reserve Bank of New York ECONOMIC POLICY REVIEW 2, no. 1 (April): 1-15.

Pritsker, Matt. 1997. "Liquidity Risk and Positive Feedback." In ON THE MEASUREMENT OF AGGREGATE MARKET RISK. Basle: Bank for International Settlements, November.

Risk. 1994. "Safer than Cash." Vol. 7, no. 5 (May): 6.

Swaps Monitor. 1996. "Returns on Capital." Vol. 10, no. 4 (December 16): 1.

U.S. Department of the Treasury. 1996. TREASURY BULLETIN, June.

Does Consumer Confidence Forecast Household Expenditure? A Sentiment Index Horse Race

Jason Bram and Sydney Ludvigson

The effect of consumer attitudes on economic activity is a subject of great interest to both policymakers and economic forecasters. Household sentiment has been cited as one of the leading causes of the 1990-91 recession,[1] and recent levels of confidence indexes have helped fuel speculation that the economy may be headed for a period of overheating. Unexpected shifts in consumer confidence have also been used to explain swings in financial markets.

Two surveys of consumer attitudes—the University of Michigan Index of Consumer Sentiment and the Conference Board Consumer Confidence Index—are widely tracked by policymakers, financial analysts, and journalists. Despite the popularity of these indexes, there is little consensus about their ability to collect information on consumer spending that is not already captured by economic fundamentals. Also uncertain is whether one survey is more informative than the other.

In response to the widespread belief that consumers' opinions and expectations influence the direction of the economy, a growing number of studies have set out to analyze the relationship between consumer attitudes and economic variables. To date, academic research has focused exclusively on the predictive power of the University of Michigan's Index of Consumer Sentiment, most likely because of its longer time series.[2] Although these studies generally do not find a significant relationship between consumer attitudes and future real economic activity, results have varied with the economic outcomes being forecast and with the indicators included as controls.[3]

The inconclusive results of the existing research on consumer attitudes leave two important questions unanswered: Does consumer sentiment provide economically meaningful information about future consumer spending beyond that already contained in other economic indicators? Is one attitudinal measure more informative than another?

Jason Bram and Sydney Ludvigson are economists at the Federal Reserve Bank of New York.

This study is the first formal investigation of consumer attitudes that compares the forecasting power of the University of Michigan's Index of Consumer Sentiment and the Conference Board's Consumer Confidence Index. We begin with a background analysis of structural differences between the Michigan and Conference Board indexes. We then undertake a formal statistical comparison of the predictive power exhibited by each overall survey and its component questions for several categories of consumer spending growth.

The University of Michigan's Consumer Sentiment Index and the Conference Board's Consumer Confidence Index are the most widely followed measures of U.S. consumer confidence.

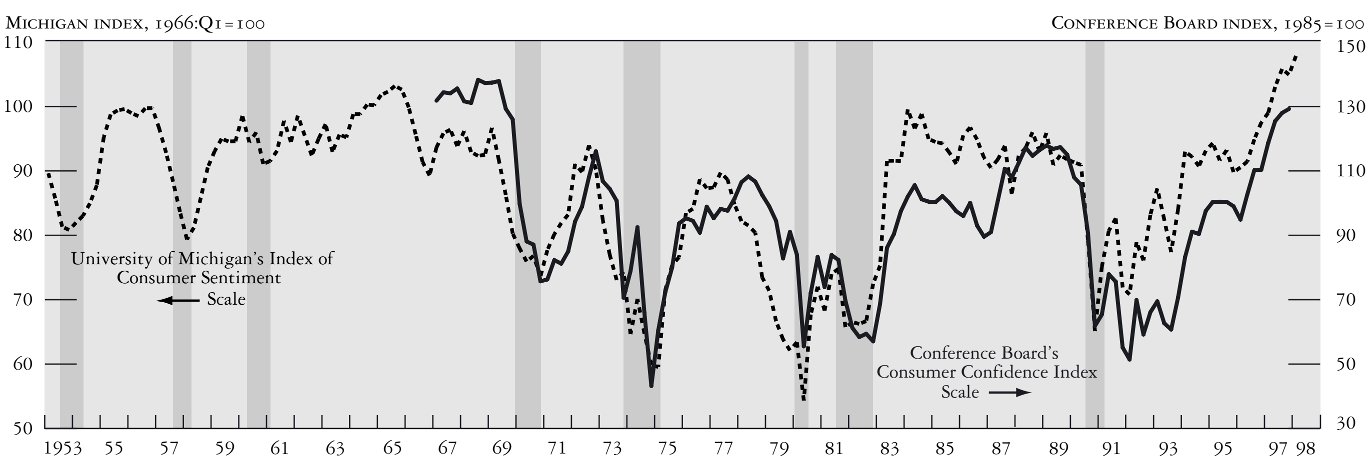
Our empirical analysis suggests that consumer sentiment can help predict future movements in consumer spending; that forecasting power, however, depends on the survey in question. Measures of consumer attitudes available from the Conference Board have both economically and statistically significant explanatory power for several spending categories—including total personal consumption expenditures; motor vehicles; services; and durables, excluding motor vehicles—even when the information contained in other economic indicators such as income, interest rates, and stock prices is known. Measures available from the University of Michigan's Survey Research Center, however, exhibit weaker forecasting power for most categories of consumer spending.[4]

A Comparison of the Michigan and Conference Board Surveys

The University of Michigan's Consumer Sentiment Index and the Conference Board's Consumer Confidence Index are the most widely followed measures of U.S. consumer confidence (Chart 1). Although the financial markets and the business community closely follow both indexes, virtually all published academic research focuses on the Michigan index—most likely because of its longer history. The Michigan index began as an annual survey in the late 1940s. In 1952, it was converted to a quarterly survey and in 1978 to a monthly survey. The Conference Board launched its index on a bimonthly basis in 1967 and expanded it to a monthly series in 1977.

Chart 1

Two Indexes of Consumer Attitudes



Sources: Conference Board; University of Michigan Survey Research Center.

Note: Shaded areas denote periods designated recessions by the National Bureau of Economic Research.

Although the two indexes broadly measure the same concept—public confidence in the economy—they are based on different sets of questions and sometimes give conflicting signals. In order to interpret movements in these two series, it is important to understand some key differences in the specific questions that are asked as well as in sample size, survey methodology, and index formulation.

Survey Questions: Present Conditions and Expectations Components

Both the Conference Board and the University of Michigan base their overall index of consumer confidence on five questions that are part of a broader survey of consumer attitudes and expectations (Box A). In addition to the overall index, both organizations report two component indexes.

Present Conditions Component

In each survey, two of the five questions ask respondents to assess present economic conditions. Michigan calls the component index based on these two questions *current conditions*, while the Conference Board uses the term *present situation*. Throughout the article, we use the generic term *present conditions* for both organizations. The present conditions questions receive a 40 percent weight in each overall index.

The Conference Board's present conditions component takes a "snapshot" approach, asking respondents to evaluate current business conditions and job availability. Because of the nature of the questions, the Conference Board's present conditions component closely tracks the nation's unemployment rate, and year-over-year changes in the index are closely correlated with payroll employment growth.

Box A: Component Questions of Consumer Confidence

Five questions make up the confidence indexes reported by the University of Michigan and the Conference Board. Each set of questions asks respondents to assess present and future economic conditions and is part of a broader monthly survey of consumer attitudes.[a]

Michigan Survey	Conference Board Survey
PRESENT CONDITIONS QUESTIONS	PRESENT CONDITIONS QUESTIONS
Q1) Do you think now is a good or bad time for people to buy major household items? [good time to buy/uncertain, depends/ bad time to buy]	Q1) How would you rate present general business conditions in your area? [good/normal/bad]
Q2) Would you say that you (and your family living there) are better off or worse off financially than you were a year ago? [better/same/worse]	Q2) What would you say about available jobs in your area right now? [plentiful/not so many/hard to get]
EXPECTATIONS QUESTIONS	EXPECTATIONS QUESTIONS
Q3) Now turning to business conditions in the country as a whole—do you think that during the next twelve months, we'll have good times financially or bad times or what? [good times/uncertain/bad times]	Q3) Six months from now, do you think business conditions in your area will be [better/same/worse]?
Q4) Looking ahead, which would you say is more likely—that in the country as a whole we'll have continuous good times during the next five years or so or that we'll have periods of widespread unemployment or depression, or what? [good times/uncertain/ bad times]	Q4) Six months from now, do you think there will be [more/same/ fewer] jobs available in your area?
Q5) Now looking ahead—do you think that a year from now, you (and your family living there) will be better off financially, or worse off, or just about the same as now? [better/same/worse]	Q5) How would you guess your total family income to be six months from now? [higher/same/lower]

[a] To compare the two indexes, we reorder the questions and number them one through five. In addition, because the University of Michigan and the Conference Board use slightly different terminology for the index component based on the first two questions, we adopt the term *present conditions* for both organizations.

Michigan asks respondents to comment on the advisability of big-ticket household purchases and to assess changes in their own financial situation. Michigan's present conditions component is less closely tied to labor market conditions and its level tends to reflect recent changes in the economy rather than the level of economic activity.

These differences are reflected in the cyclical behavior of the two present conditions component indexes: Michigan's generally peaks in the early stages of economic recovery, when growth is high. By contrast, the Conference Board's generally peaks in the late stages of economic expansion, when unemployment is low and the level of economic activity is high. Not surprisingly, given the differences in the questions, the present conditions components of the two indexes are not closely correlated (Chart 2).

Expectations Component

The three questions that ask about consumers' expectations are fairly comparable in the two surveys. The Conference Board survey asks about expected changes in business conditions, job availability, and respondents' income over the next six months.[5] Michigan's poses questions on expected business conditions—both over the next year and over the next five years—and expected changes in the respondent's financial situation over the next year.[6]

Unlike the present conditions components, the expectations components in the two surveys are highly correlated with each other (Chart 3). Moreover, Michigan's present conditions and expectations components are much more closely correlated than are the Conference Board's (Appendix A).

Methodology

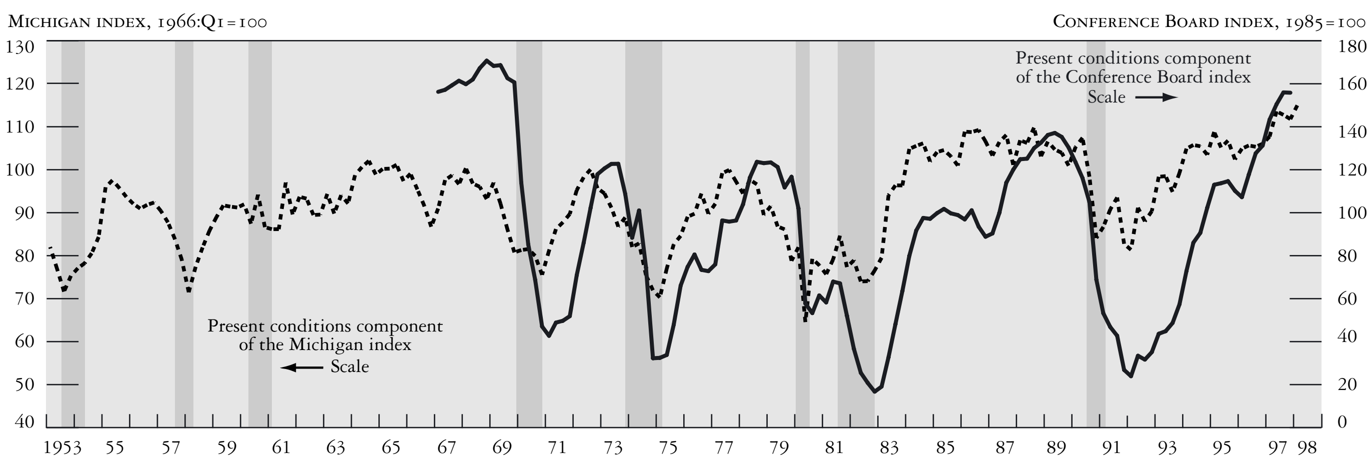
The most important methodological differences between the two surveys concern sample size, which affects sampling error and thus reliability, and index construction, which affects the range of movement in the indexes. The survey timing and release schedules also differ—a relevant consideration when conducting real-time analysis.

Michigan conducts its survey by phone throughout most of the month. Its sample size is 500; a preliminary midmonth release is based on about 250 phone interviews conducted early in the month. Final figures for the full sample are subsequently made available at the end of the month and are not subject to further revision.

The Conference Board sends out a mail survey at the end of the prior month and responses flow in throughout the survey month. The sample size is roughly 3,500 (of a total mailing of 5,000).[7] On the last Tuesday of the survey month, the Conference Board formally releases its preliminary figures

Chart 2

PRESENT CONDITIONS COMPONENT OF CONSUMER ATTITUDES

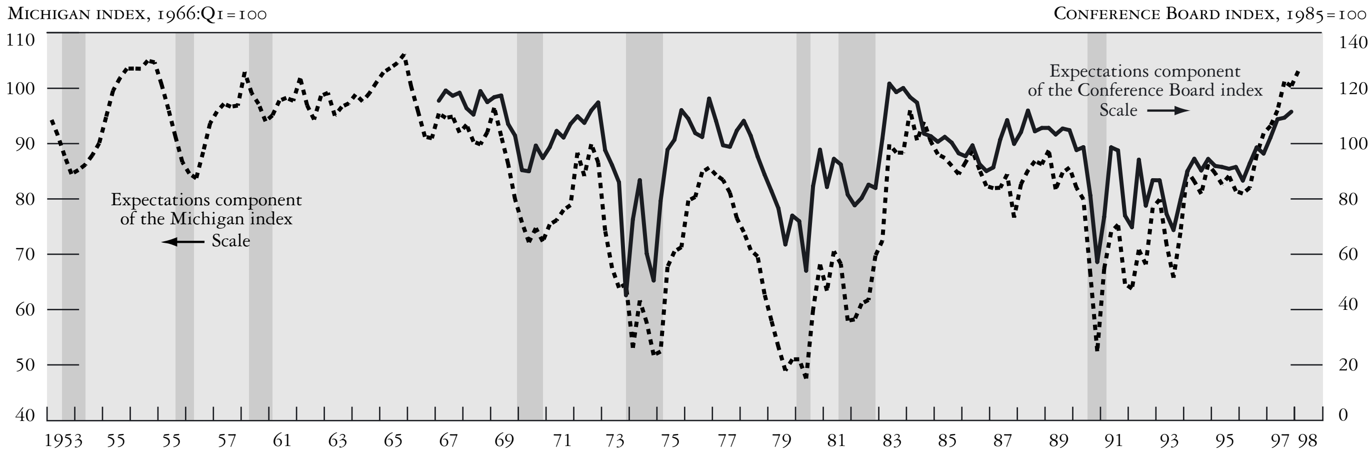


Sources: Conference Board; University of Michigan Survey Research Center.

Note: Shaded areas denote periods designated recessions by the National Bureau of Economic Research.

Chart 3

EXPECTATIONS COMPONENT OF CONSUMER ATTITUDES



Sources: Conference Board; University of Michigan Survey Research Center.

Note: Shaded areas denote periods designated recessions by the National Bureau of Economic Research.

based on about 2,500 responses. Final, revised data based on the full monthly sample are released with the next month's preliminary figures and are not subject to further revision.

The University of Michigan and the Conference Board also use different methodologies to construct their indexes from the raw response data (Box B). The main result of these methodological differences is that the Conference Board's overall index and component measures have a wider range of movement than Michigan's. In other words, identical shifts in the underlying responses tend to produce significantly larger moves in the Conference Board's indexes than in Michigan's.

A good rule of thumb is that a one-point move in Michigan's index is roughly comparable to a two-point move in the Conference Board's Index.

Interpretation of the Indexes

Although the Conference Board and Michigan indexes are highly correlated, they sometimes move independently of one another. Because of differences in survey methodology, index construction, and base year, index levels are not comparable; monthly changes must be compared on a standardized basis rather than in absolute terms. A good rule of thumb is that a one-point move in Michigan's index is roughly comparable to a two-point move in the Conference Board's index.

The indexes also differ in timeliness and reliability. One advantage of Michigan's index is that its preliminary figures are available earlier than the Conference Board's. However, because Michigan's figures are based on a much smaller sample size than the Conference Board's, they are more susceptible to measurement error. As a result, random monthly fluctuations tend to be more pronounced in Michigan's index than in the Conference Board's.[8]

Two of the most common dilemmas in relying on consumer confidence as an economic indicator are whether to focus on index level or month-to-month changes and whether to focus on the present conditions or the expectations component. For the Conference Board index, it is particularly useful to examine the present conditions and expectations components individually. The level of the present conditions component serves as a good proxy for the level of economic activity, while the expectations com-

BOX B: CALCULATING THE TWO INDEXES

The example below illustrates how the Conference Board and Michigan would construct a single index for one question using the same raw response data. Hypothetical figures are shown for two months along with the base-period levels against which the indexes are benchmarked.

Michigan calculates a diffusion measure by adding the difference between the positive and negative percentages to 100. Thus, the current month's value is 112 $[100 + 24 - 12]$, and the prior month's level is 120 $[100 + 30 - 10]$. Next, an index is constructed by dividing the level of the diffusion measure by the base-period level of 110, and then multiplying by 100. In the example below, this calculation yields a value of 101.8 $[120 + 110 - 100]$ for the current month, down from the prior month's level of 109.1 $[120 + 110 - 100]$ — a drop of 7.3 points.

EXAMPLE: CALCULATING INDEX LEVELS FROM RAW RESPONSE DATA

	Base Period	Prior Month	Current Month
Percentage of responses			
Positive	25	30	24
Neutral	60	60	64
Negative	15	10	12
Indicator level			
Michigan diffusion measure	110.0	120.0	112.0
Michigan index	100.0	109.1	101.8
Conference Board diffusion measure	62.5	75.0	66.7
Conference Board index	100.0	120.0	106.7

Using the same raw responses, the Conference Board would calculate its diffusion measure by dividing the positive response percentage by the sum of the positive and negative response percentages. This procedure gives a value of 66.7 $[24 \div (24 + 12) \times 100]$ for the current month and 75 $[30 \div (30 \times 10) \times 100]$ for the prior month. Next, the index is calculated to be 106.7 $[66.7 \div 62.5 \times 100]$ in the current month, down from a level of 120 $[75.0 \div 62.5 \times 100]$ in the prior month—a drop of 13.3 points.

Some subtle differences in index construction are not illustrated here. First, the Conference Board converts each diffusion index to a base-year index and then averages the indexes together.[a] Michigan first averages the diffusion indexes into a composite diffusion index and then converts the results to a base-period index. Second, the Conference Board's responses are seasonally adjusted, while Michigan's are not. However, the seasonal adjustment has little effect on our results, because neither index exhibits much seasonality. Finally, because the Conference Board and Michigan use different base periods (1985 and 1966:Q1, respectively), the response patterns on which the indexes are based may differ. As a result, the index *levels* of the two surveys are not comparable.

[a] Because the Conference Board's diffusion measures are converted into base-year indexes before they are averaged arithmetically, a given question's effective weight in the index is influenced by the selection of the base year. In theory, the choice of the base year could affect the magnitude and even the direction of change in the index. (The resulting problems are similar to those associated with the old fixed-base-year GDP deflator.) In practice, however, this feature has no discernible effect on the Conference Board's index.

ponent is more closely correlated with the rate of economic growth. In Michigan's survey, both components are closely correlated and in general serve as an indicator of the pace of economic growth.

The Conference Board index, the Michigan index, and the components of each index exhibit some movement that cannot be explained by movements in other economic indicators such as income, interest rates, and lagged consumption. In the next section, we determine whether this independent movement contains information that can help predict consumer spending.

EMPIRICAL RESULTS

We use a two-step procedure to determine the forecasting power of consumer confidence. First, we consider a baseline forecasting equation for consumption growth that does not include attitudinal survey measures. We then add consumer sentiment to the baseline equation and test which measures of consumer attitudes, if any, improve the forecasting power of the baseline equation. In estimating the confidence-augmented equation, we employ two types of statistical tests to determine whether consumer attitudes help predict future movements in consumer spending: in-sample regressions and out-of-sample regressions of consumption growth. The in-sample tests investigate the predictive power of consumer sentiment over the entire sample period; the out-of-sample procedure tests the stability of that predictive power over several subsamples of the data.

Our analysis measures the effect of consumer attitudes on five categories of household personal consumption expenditure: total expenditure; motor vehicle expenditure; expenditure on all goods, excluding motor vehicles; expenditure on services; and expenditure on durable goods, excluding motor vehicles. The data are quarterly and span the period from the first quarter of 1967 to the third quarter of 1996.[9] Definitions of the variables used in the equations appear in Appendix B.

Baseline Forecasting Equation

We specify a simple forecasting equation for consumption growth that does not include consumer confidence. This specification, or the baseline equation, takes the form

$$(1) \qquad \Delta \ln(C_t) = \alpha_0 + \gamma Z_{t-1} + \varepsilon_t ,$$

where C_t is real consumption spending and Z_{t-1} is a vector of control variables. In choosing economic indicators to include in Z_{t-1}, we adhere to the existing literature closely. In an earlier work, Carroll, Fuhrer, and Wilcox (1994) estimated a similar equation to test whether the Michigan index contained any incremental predictive power for future movements in consumer spending. Their baseline equation placed lagged values of the dependent variable and of labor income growth in Z. The inclusion of labor income growth as a control variable is motivated by a large and growing body of empirical work showing that consumption growth is related to lagged, or predictable, income growth (see, for example, Flavin [1981] and Campbell and Mankiw [1989]). Like Carroll, Fuhrer, and Wilcox, we include these indicators on the right-hand-side of the equation using four lags of each variable. As is typical in aggregate time series, Akaike and Schwarz tests did not indicate the need for more than four quarterly lags.

Other researchers have argued that the information contained in attitudinal indicators should be assessed relative to that contained in financial indicators. Leeper (1992) points out that consumer sentiment may have predictive power for spending because consumer surveys are made available on a more timely basis than other economic indicators such as income and consumption data. However, he goes on to argue that financial market indicators are available on an almost continuous basis and may contain much of the same information captured by consumer sentiment. Indeed, Leeper finds that consumer attitudes are only weakly correlated with variables such as unemployment and industrial production once financial indicators are included. To investigate whether consumer attitudes contain useful information for future consumer spending beyond that contained in financial indicators, we follow Leeper's suggestion and include the log first difference of the real stock price and the first difference of the three-month Treasury bill rate in our Z vector.

Our analysis measures the effect of consumer attitudes on five categories of household personal consumption expenditure: total expenditure; motor vehicle expenditure; expenditure on all goods, excluding motor vehicles; expenditure on services; and expenditure on durable goods, excluding motor vehicles.

As a robustness check for our Z specification, we substitute the unemployment rate for labor income growth. In addition, we substitute three different variables for the first difference of the three-month Treasury bill rate: the spread between the ten-year Treasury bond rate and the one-year Treasury bill rate,[10] the first difference of the one-year Treasury bill rate, and the first difference of the federal funds rate. The results, which are not reported here, indicate that these substitutions do not qualitatively alter the estimation of the baseline model. To summarize, the control variables included in Z are four lags of the dependent variable, four lags of the growth in real labor income, four lags of the log first difference in the real stock price index as measured by the Standard and Poor's 500 index, and four lags of the first difference of the three-month Treasury bill rate.

According to our estimation, lagged values of consumption growth and the financial indicators in Z have predictive power for most categories of consumer expenditure. Table 1 presents the estimation results of the baseline model. For each category of consumption, the table presents the sum of the coefficients on the lags of each variable in Z. The sum of the coefficients on the four lags of each variable estimates the long-run effect of the variable on consumption growth. The *p*-values for the joint marginal significance of the lags of each variable, which appear in parentheses, give the probability that the explanatory variable can be excluded from the forecasting equation.[11] When the *p*-values are very low, the variables are statistically significant predictors of consumption growth.

As Table 1 shows, the long-run impact of most variables has the expected sign. Consumption growth is positively related to lagged consumption growth for most of the categories, while lagged interest rates have a small negative effect on future consumption. Interestingly, the inclusion of the consumption and interest rate variables appears to reduce the statistical significance of the income and stock market variables in forecasting consumption growth.

Table 1
BASELINE FORECAST OF CONSUMPTION GROWTH

Predicted Variable	Four Lags of Consumption	Four Lags of Income	Four Lags of Treasury Bill Rate	Four Lags of S&P 500
Total	0.83	0.04	-0.002	-0.01
	(0.000)	(0.263)	(0.006)	(0.056)
Motor vehicles	0.47	0.40	-0.024	-0.05
	(0.230)	(0.221)	(0.068)	(0.012)
Goods, excluding motor vehicles	0.88	0.04	-0.001	0.0
	(0.000)	(0.356)	(0.094)	(0.148)
Services	0.05	0.50	-0.007	-0.02
	(0.021)	(0.102)	(0.000)	(0.276)
Durables goods, excluding motor vehicles	0.80	0.16	-0.006	0.0
	(0.000)	(0.886)	(0.013)	(0.477)

Source: Authors' calculations.

Notes: The table reports the sum of the coefficients on the lags of the variable indicated; the probability that the variable can be excluded from the prediction equation appears in parentheses. Hypothesis tests were conducted using a heteroskedasticity and serial correlation robust covariance matrix. The sample covers the period from the first quarter of 1968 to the third quarter of 1996. S&P=Standard and Poor's.

ADDING CONSUMER CONFIDENCE TO THE BASELINE EQUATION

To determine whether consumer attitudes help forecast future consumer spending, we add a measure of consumer confidence to the baseline equation:

$$(2) \qquad \Delta \ln(C_t) = \alpha_0 + \sum_{i=1}^{n} \beta_i S_{t-i} + \gamma Z_{t-1} + \varepsilon_t ,$$

where S is consumer confidence as measured by either the Michigan or the Conference Board index. We then replace the overall index with the expectations component as our measure of S.[12]

Our modified equation attempts to quantify the power of each index to predict future consumption expenditures. In our estimations, we report the *increment* to the adjusted R^2 that results from augmenting the baseline equation to include each of the attitudinal indicators. For example, if the increment to the adjusted R^2 from adding the four lags of S is X percent, the confidence-augmented equation predicts about X percent more of the variation in the next quarter's consumption than the baseline equation.

The first two columns of Table 2 present the results of estimating the confidence-augmented equation. The first column of Table 2 reports the results for the equation that includes the Michigan overall index (rows 1-5) and its expectations component (rows 6-10); the second column of Table 2 reports the results for the equation that includes the Conference Board overall index (rows 1-5) and its expectations component (rows 6-10). The probability that the confidence variables can be excluded from the forecasting equation appears in parentheses.[13]

Our results reveal a gap in the indexes' forecasting power for total personal consumption growth. For the Michigan survey, the lagged values of consumer sentiment do not increase the adjusted R^2 in the regression where total personal consumption growth is the dependent variable. Indeed, the inclusion of Michigan's overall index actually weakens the predictive power of the baseline equation. We obtain similar results using the Michigan expectations component. By contrast, both the Conference Board's overall measure of consumer confidence and its measure of consumer expectations are incrementally informative about the future path of total personal consumer spending growth. Adding the last four quarters of data from the Conference Board's overall confidence index to the baseline equation predicts an additional 9 percent of the variation in the next period's consumption growth. Similarly, adding the last four quarters of data on the expectations component predicts an additional 12 percent of the variation in future consumer spending. Moreover, the Conference Board index is statistically significant at better than the 5 percent level.[14]

For motor vehicle spending, however, both overall indexes display some incremental predictive power. Lagged values of the Michigan sentiment index explain an additional 5 percent of the growth in motor vehicle spending, a relatively small amount, although the increase is statistically significant at the 6 percent level. By including lags of the Michigan expectations component, however, we increase the fraction of regression variance explained by consumer confidence to 8 percent, and the expectations variables become significant at the 5 percent level. The Conference Board measures have an equal or somewhat stronger incremental impact on growth in motor vehicle spending; the overall index, like the Michigan index, increases the adjusted R^2 by 5 percent, but the inclusion of four lags of the the Conference Board's expectations component increases the adjusted R^2 by 10 percent.[15]

Adding the last four quarters of data from the Conference Board's overall confidence index to the baseline equation predicts an additional 9 percent of the varation in the next period's consumption growth.

For spending on services and durable goods (excluding motor vehicles), lagged values of either Michigan's overall index or its expectations component generally add little or no explanatory power to the consumption growth regressions. For services spending growth, the incremental adjusted R^2 is negative. The Michigan index does help to forecast growth in the goods (excluding motor vehicles) category. Still, even in this case, the inclusion of four lags of Michigan's overall index improves the forecasting performance of the baseline equation by just 3 percent.

Table 2
FORECAST OF CONSUMPTION GROWTH, AUGMENTED BY CONSUMER CONFIDENCE INDICATORS

Real Personal Consumption Expenditures	Michigan Index	Conference Board Index	Both
	Overall Index		
Total	-0.04	0.09	0.13
Conference Board	—	(0.001)	(0.000)
Michigan	(0.715)	—	(0.040)
Motor vehicles	0.05	0.05	0.21
Conference Board	—	(0.020)	(0.000)
Michigan	(0.059)	—	(0.000)
Goods, excluding motor vehicles	0.03	0.07	0.05
Conference Board	—	(0.177)	(0.392)
Michigan	(0.000)	—	(0.934)
Services	-0.02	0.02	0.11
Conference Board	—	(0.062)	(0.001)
Michigan	(0.607)	—	(0.140)
Durables, excluding motor vehicles	0.00	0.15	0.17
Conference Board	—	(0.005)	(0.041)
Michigan	(0.257)	—	(0.780)
	Expectations Component		
Total	-0.03	0.12	0.11
Conference Board	—	(0.000)	(0.000)
Michigan	(0.557)	—	(0.645)
Motor vehicles	0.08	0.10	0.19
Conference Board	—	(0.006)	(0.000)
Michigan	(0.042)	—	(0.014)
Goods, excluding motor vehicles	0.00	-0.12	-0.02
Conference Board	—	(0.334)	(0.696)
Michigan	(0.858)	—	(0.884)
Services	-0.01	0.06	0.07
Conference Board	—	(0.018)	(0.010)
Michigan	(0.554)	—	(0.253)
Durables, excluding motor vehicles	0.03	0.06	0.02
Conference Board	—	(0.217)	(0.677)
Michigan	(0.298)	—	(0.687)

Source: Authors' calculations.

Notes: The table reports the increment to the adjusted R^2 statistic from adding four lags of the confidence measures; *p*-values for the joint marginal significance of the lags of the confidence measures appear in parentheses. Hypothesis tests were conducted using a heteroskedasticity and serial correlation robust covariance matrix. The sample covers the period from the first quarter of 1968 to the third quarter of 1996.

Lagged values of the Conference Board's overall index appear to be of value in predicting spending in durables and services. For durable goods (excluding motor vehicles), adding lags of the overall Conference Board index increases the fraction of regression variance explained by consumer confidence by 15 percent, a finding that is highly statistically significant. Moreover, the Conference Board expectations component appears strongly related to future services expenditures—the addition of that variable increases the adjusted R^2 by a statistically significant 6 percent.

For the category of motor vehicle expenditure, including both measures of consumer attitudes in the forecasting equation may be superior to the use of either index alone.

HEAD-TO-HEAD FORECASTING COMPETITION

The results discussed above suggest that the Conference Board index generally serves as a better predictor of spending than the Michigan index. Despite our finding that the Michigan index has little explanatory power for categories of expenditure other than motor vehicles, it is still possible that the Conference Board index contains only a small amount of information that is independent of that contained in the Michigan index. If this were true, including both indexes in the equation simultaneously could substantially reduce the explanatory power of the Conference Board index. To examine this possibility, we estimate a "head-to-head" forecasting equation that includes both measures of consumer attitudes in the equation at the same time and takes the form

$$(3) \qquad \Delta \ln(C_t) = \alpha_0 + \sum_{i=1}^{4} \beta_i S_{t-i}^{c} + \sum_{i=1}^{4} \delta_i S_{t-i}^{m} + \gamma Z_{t-1} + \varepsilon_t ,$$

where S^c and S^m are the consumer confidence variables as measured by the Conference Board index and the Michigan index, respectively. As in equation 2, we report results when S^c and S^m are measured as each survey's overall index or its expectations component.

The results of estimating equation 3 appear in the third column of Table 2. The numbers reported for both indexes are the increment to the adjusted R^2 after both confidence measures are added to the baseline equation. The probability that the Conference Board and Michigan indexes can be excluded from equation 3 appears in parentheses. The table shows that the Conference Board variables remain statistically significant once the Michigan variables are included. Thus, the direct inclusion of both the Michigan index and the Conference Board index in the forecasting equation does not eliminate the forecasting power of the Conference Board index. Indeed, for the category of motor vehicle expenditure, including both measures of consumer attitudes in the forecasting equation may be superior to the use of either index alone. The increment to the adjusted R^2 from adding both overall sentiment measures to the motor vehicle baseline regression is 21 percent, a large increase over that obtained when the equation incorporates only one of the indexes. Moreover, both indexes remain statistically significant predictors of motor vehicle expenditure in the head-to-head specification. We discuss one possible explanation for this finding in the question-level analysis below.

As we have shown, the results in this section suggest that a gap exists in the predictive power of the two attitudinal surveys, with the Conference Board index generally outperforming the Michigan index. We now examine whether this gap can be explained by differences in the individual questionnaires.

QUESTION-LEVEL ANALYSIS

The underlying questions of the Conference Board and Michigan indexes serve as mini-diffusion indexes that are similar in construction to the overall indexes.[16] We test the predictive power of each question-level index using the following equation:

$$(4) \qquad \Delta \ln(C_t) = \alpha_0 + \sum_{i=1}^{4} \beta_i Q_{jt-i}^{k} + \gamma Z_{t-1} + \varepsilon_t ,$$

where Q_j^k denotes question j of index k, for j=1,...5 and k=1,2.

As we would expect, no single question helps predict all categories of spending growth. Several questions, however, help predict growth in particular categories of expenditure. Table 3 reports the increment to the adjusted R^2 from adding four lags of each question to the baseline equation. As the table shows, questions two and four of the Conference Board index explain a substantial portion of the regression variance for total consumption, motor vehicle, and services expenditures (up to 14 percent for the motor vehicle category). Moreover, the Conference Board's question one is a strong predictor of durable goods (excluding motor vehicles) spending, yielding an incremental adjusted R^2 of 18 percent. In addition, for both indexes, questions three, four, and five hold predictive power for motor vehicle expenditures. The Conference Board's questions three through five also help explain total expenditures.[17]

From the results in Table 3, we arrive at several general conclusions about the types of questions that have significant forecasting ability. First, questions that ask specifically about job prospects in the respondent's area (questions two and four of the Conference Board survey) generally have the most explanatory power.

Second, questions that ask about either the present or the future have more forecasting power than questions that compare the present with the past. The Michigan index's question two, the only question in either index that asks about conditions today relative to the past, has virtually no explanatory power.

Third, questions that ask about consumers' personal financial situations exhibit more predictive power than questions that ask about present buying conditions: for both surveys, the question on personal finances (question five) is significant for some categories of expenditure. The only question about current buying conditions, question one of the Michigan index, elicits virtually no incremental information.

Table 3

PREDICTIVE POWER OF THE SURVEYS' COMPONENT QUESTIONS

Real Personal Consumption Expenditures	Michigan Index		Conference Board Index	
	Present Conditions Component		Present Conditions Component	
	Question 1	Question 2	Question 1	Question 2
Total	-0.01	-0.03	0.04	0.10
	(0.542)	(0.482)	(0.037)	(0.002)
Motor vehicles	0.06	0.02	0.02	0.14
	(0.066)	(0.191)	(0.023)	(0.001)
Goods, excluding motor vehicles	0.09	0.02	0.05	0.07
	(0.128)	(0.262)	(0.337)	(0.200)
Services	-0.01	0.03	0.06	0.06
	(0.322)	(0.149)	(0.035)	(0.058)
Durables goods, excluding motor vehicles	0.03	0.05	0.18	0.13
	(0.406)	(0.086)	(0.002)	(0.004)

	Expectations Component			Expectations Component		
	Question 3	Question 4	Question 5	Question 3	Question 4	Question 5
Total	-0.03	-0.03	0.02	0.09	0.09	0.06
	(0.488)	(0.395)	(0.179)	(0.001)	(0.001)	(0.012)
Motor vehicles	0.04	0.05	0.06	0.10	0.10	0.03
	(0.107)	(0.040)	(0.017)	(0.002)	(0.003)	(0.115)
Goods, excluding motor vehicles	0.01	0.02	0.03	0.0	0.0	0.03
	(0.675)	(0.832)	(0.407)	(0.506)	(0.762)	(0.105)
Services	-0.02	0.01	-0.02	0.01	0.04	0.08
	(0.617)	(0.339)	(0.502)	(0.470)	(0.087)	(0.010)
Durable goods, excluding motor vehicles	0.02	0.06	0.03	0.0	0.06	0.03
	(0.266)	(0.188)	(0.117)	(0.451)	(0.162)	(0.289)

Source: Authors' calculations.

Notes: Figures in parentheses are *p*-values of the joint significance of the lags of the component question. Hypothesis tests were conducted using a heteroskedasticity and serial correlation robust covariance matrix. The sample covers the period from the first quarter of 1968 to the third quarter of 1996.

Fourth, consumer expectations over long-term horizons may be more informative than expectations over short-term horizons for predicting expenditures on large-ticket items such as motor vehicles.[18] Note that questions three, four, and five of the Michigan index are positively correlated with future spending on motor vehicles: the coefficients on the four lags of question three are jointly significant at the 10 percent level, and the coefficients on the four lags of questions four and five, at better than the 5 percent level. These three questions ask about consumers' expectations over a time horizon of one year or more.

Finally, the differences in the types of questions asked by the two surveys may explain our earlier finding that including both measures of consumer attitudes in the forecasting equation better predicts motor vehicle spending than does the use of either index alone. Using both surveys allows the model to capture simultaneously two aspects of consumer sentiment that appear important to motor vehicle spending: consumer expectations over long-term horizons (the Michigan survey) and consumer expectations about job availability (the Conference Board survey).

Out-of-Sample Tests

Our results so far have been obtained by estimating the confidence-augmented equations over the whole sample period. In this section, we test the ability of the equation to forecast out of sample. These tests indicate that the out-of-sample predictive power of the overall Conference Board index was strong in the 1980s, but that it diminished in the early 1990s. Out-of-sample forecast equations augmented with the Michigan index do not generally improve upon the predictive power of the baseline model in any subperiod.

Out-of-Sample Forecasting Performance

To conduct the out-of-sample forecasts, we compare the forecast accuracy of equation 2 over two nonoverlapping evaluation periods across specifications that include either the overall index of each survey or one of the survey's component questions.

The out-of-sample procedure is as follows: as before, the baseline model specifies consumption growth as a function of four lags of the dependent variable, four lags of the growth in real labor income, four lags of the log first difference in the real stock price index as measured by the Standard and Poor's 500 index, and four lags of the first difference of the three-month Treasury bill rate. We then analyze the out-of-sample forecast error of the confidence-augmented models. The model is first estimated using data from the first quarter of 1968 to the fourth quarter of 1981. We then conduct out-of-sample forecasts for two subperiods: the first quarter of 1982 to the fourth quarter of 1989 and the first quarter of 1990 to the third quarter of 1996. We use recursive regressions to reestimate the model, adding one quarter at a time and calculating a series of one-step-ahead forecasts. The forecasts are evaluated by computing the root-mean-squared error from the set of one-step-ahead forecasts.

> *{Out-of-sample} tests indicate that the . . . predictive power of the overall Conference Board index was strong in the 1980s, but that it diminished in the early 1990s.*

Chart 4 provides a visual impression of the relative forecasting power of each model for four different periods. The chart compares the implied consumption levels of the models using the overall indexes with actual levels obtained during those years. As the chart shows, during several episodes in the 1980s and 1990s, the Conference Board-augmented model predicts a level of consumption that was closer to the actual level than that predicted by either the baseline or the Michigan-augmented models.

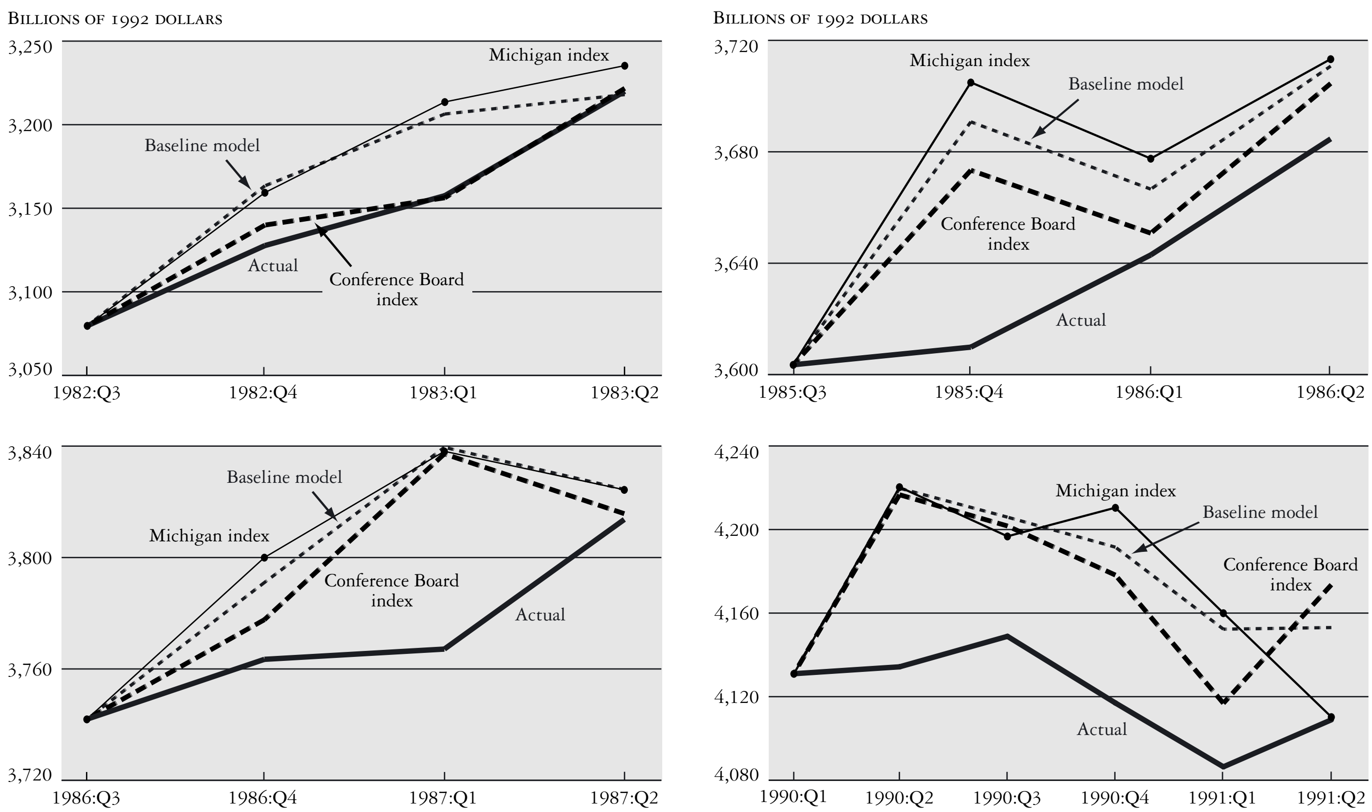
Table 4 summarizes the out-of-sample forecasting performance of each confidence-augmented model. We compare the accuracy of the confidence-augmented equations with that of the baseline model. For each evaluation period and each category of consumer expenditure, the first entry gives the ratio of the root-mean-squared error obtained for the Michigan-augmented model to that obtained for the baseline model.

The second entry gives the ratio of the root-mean-squared error for the Conference Board-augmented model to that obtained for the baseline model. In both cases, results of less than one indicate that using the attitudinal indicator in the forecasting equation improves the out-of-sample forecast relative to the baseline equation. Finally, the third entry gives the ratio of the root-mean-squared error of the Michigan-augmented model to that of the Conference Board-augmented model; a number greater than one indicates that the Conference Board-augmented model outperformed the Michigan-augmented model. The modified Diebold-Mariano test statistic for equal forecast accuracy (see Harvey, Leybourne, and Newbold [1997]) appears in parentheses. We discuss the use of this test statistic below.

For most categories of consumer spending, the forecasting error of the Conference Board-augmented equation is lower than that of the Michigan-augmented equation over most evaluation periods. Moreover, for total personal consumption expenditures and motor vehicle expenditures, the forecasting error of the Conference Board-augmented equation is lower than that of the baseline equation during the 1980s. For example, for total personal consumption expenditures, inclusion of the Conference Board index reduces the root-mean-squared error over the baseline equation by 10 percent for the period from first-quarter 1982 to third-quarter 1996, and by 14 percent for the subperiod from first-quarter 1982 to fourth-quarter 1989. By contrast, the Michigan-augmented equation performs worse than the

Chart 4

Implied Consumption Levels: Actual Relative to Forecast



Source: Authors' calculations.

Note: Dollars are chain-weighted.

baseline model in predicting growth in total personal consumption spending during both of these periods. A similar result holds for the equations that predict growth in motor vehicle spending.[19]

Although the out-of-sample results in Table 4 reveal many qualitative similarities with the in-sample results, a few differences in outcome arise when we estimate the equations over different subperiods. Most notably, while the Michigan index is found to be helpful in forecasting future movements in motor vehicle expenditures when the equation is estimated over the full sample, the out-of-sample results reveal that including the Michigan index improves the predictive power of the baseline equation only in the subperiod from first-quarter 1990 to third-quarter 1996 and weakens the forecasts over the entire first-quarter 1982 to third-quarter 1996 period. Moreover, the out-of-sample predictive power of the Conference Board index appears to be concentrated in the total personal consumption category and in motor vehicle spending. In contrast to the strong in-sample predictive power displayed in Table 2, the Conference Board model does not improve the forecasting performance of the baseline equation in any subperiod for expenditures on goods (excluding motor vehicles).

Table 4
OUT-OF-SAMPLE PREDICTIVE POWER OF ONE-STEP-AHEAD FORECASTS

Real Personal Consumption Expenditures	1982:Q1–1996:Q3	1982:Q1–1989:Q4	1990:Q1–1996:Q3
Total			
Michigan/baseline model	1.014	1.035	1.037
	(0.68)	(0.60)	(0.38)
Conference Board/baseline model	0.900	0.857	1.042
	(-0.70)	(-1.40)	(0.35)
Michigan/Conference Board	1.127	1.208	0.995
	(1.14)	(1.42)	(-0.06)
Motor vehicles			
Michigan/baseline model	1.019	1.029	0.998
	(0.41)	(0.64)	(-0.02)
Conference Board/baseline model	0.930	0.902	0.988
	(-1.17)	(-1.50)	(-0.10)
Michigan/Conference Board	1.096	1.141	1.010
	(1.56)	(1.70)*	(0.11)
Goods, excluding motor vehicles			
Michigan/baseline model	0.990	0.994	1.035
	(-0.25)	(-0.91)	(0.58)
Conference Board/baseline model	1.016	1.013	1.020
	(0.37)	(0.22)	(0.28)
Michigan/Conference Board	0.974	0.981	1.014
	(-0.52)	(-0.80)	(0.25)
Services			
Michigan/baseline model	1.081	1.125	1.030
	(1.87)*	(1.82)*	(0.62)
Conference Board/baseline model	1.029	1.004	1.056
	(0.47)	(0.06)	(0.53)
Michigan/Conference Board	1.051	1.121	0.975
	(0.92)	(1.65)	(-0.30)
Durable goods, excluding motor vehicles			
Michigan/baseline model	1.040	1.024	1.075
	(1.19)	(0.59)	(1.19)
Conference Board/baseline model	1.061	1.088	0.996
	(1.48)	(1.66)	(-0.09)
Michigan/Conference Board	0.980	0.941	1.079
	(-0.41)	(-1.01)	(1.18)

Source: Authors' calculations.

Notes: The table reports the ratio of the root-mean-squared forecasting error. A number less than one indicates that the confidence-augmented model in the numerator has superior forecasting ability. The modified Diebold-Mariano test statistics (Harvey, Leybourne, and Newbold 1997) appear in parentheses. Out-of-sample evaluation periods are reported at the top of each column; the initial estimation period begins with the first quarter of 1968 and ends with the fourth quarter of 1981.

*Significant at the 10 percent level.

The numbers in parentheses in Table 4 give the modified Diebold-Mariano test statistic derived from the method in Harvey, Leybourne, and Newbold (1997) for testing equal forecast accuracy. This statistic has a student's t-distribution and allows the researcher to test whether differences in root-mean-squared error are statistically significant. For each category of consumer expenditure, the statistics indicate whether the out-of-sample forecast error of the confidence-augmented equation is statistically greater than the forecast error of the baseline equation. A positive number indicates that the baseline model has a lower forecast error than the confidence-augmented model. The forecast errors of the confidence-augmented models are also compared with one another; a positive test statistic indicates that the Conference Board-augmented model has a lower forecast error than the Michigan-augmented model.

We report these test statistics but remain skeptical about their value in detecting differences in forecast accuracy. A number of recent papers have documented problems with procedures that test whether differences in out-of-sample forecast error are statistically significant. Researchers often find that variable x Granger-causes variable y in sample, but that out-of-sample tests detect no statistically significant difference in forecast accuracy across the two models according to whether or not they include x. One possible explanation for differences in in-sample and out-of-sample forecast accuracy is that the in-sample procedure may over-fit the data relative to the out-of-sample procedure. A second possible explanation is that out-of-sample tests simply have little power to reject the null hypothesis of equal forecast accuracy. Clark (1996) shows that tests for equal out-of-sample forecast accuracy generally have much lower power than in-sample Granger causality tests. Thus, the Clark study demonstrates that the discrepancy between in-sample and out-of-sample results may often be attributable to the low power of tests for equal out-of-sample forecast accuracy rather than to true over-fitting in sample. This may explain why we find strong in-sample Granger causality using the Conference Board index and generally no statistically significant difference in the out-of-sample forecasting performance of our models.[20]

Not surprisingly, the test statistics in Table 4 reveal no statistically significant differences in forecast error between the baseline model and the confidence-augmented models for most categories of consumption expenditure over most evaluation periods.

In summary, the results in Table 4 indicate that using the Conference Board index of consumer confidence would have consistently improved out-of-sample forecasts of total or motor vehicle spending growth in the 1980s. After 1990, however, the forecasting power of the model appears to change (Table 4, column 3). In predicting all categories of spending growth except motor vehicles, the baseline model outperforms both confidence-augmented models during this subperiod. Whether the Conference Board index will prove a reliable predictor of consumer spending in the future remains an open question. It is too early to tell whether the forecasting power of consumer confidence displayed by the Conference Board's overall index in the 1980s will return in the late 1990s.

Question-Level Analysis

As a last step, we analyze the out-of-sample forecasting performance of each question over each evaluation period and for every category of expenditure. Because of the large number of results, we present only those combinations for which at least one of the question-level indexes displayed modest improvement in the forecasting power over the baseline model (Table 5).

As Table 5 shows, the best results over the entire period from first-quarter 1982 to third-quarter 1996 are for the confidence-augmented model that uses four lags of the Conference Board's question four on future job avail-

Table 5
OUT-OF-SAMPLE PREDICTIVE POWER OF ONE-STEP-AHEAD FORECASTS

	1982:Q1–1996:Q3		1982:Q1–1989:Q4		1990:Q1–1996:Q3	
Real Personal Consumption Expenditures	Michigan	Conference Board	Michigan	Conference Board	Michigan	Conference Board
Total						
Question 1	1.074	0.985	1.117	0.984	1.018	0.987
Question 2	1.002	1.041	1.022	1.069	0.977	1.006
Question 3	0.996	0.955	0.983	0.907	1.011	1.012
Question 4	0.989	0.916	0.980	0.846	1.000	0.995
Question 5	1.006	0.999	1.037	0.913	0.965	1.095
Motor vehicles						
Question 1	0.959	0.957	0.982	0.938	0.908	0.999
Question 2	1.005	0.977	1.039	0.994	0.926	0.940
Question 3	1.016	0.944	1.012	0.918	1.024	1.000
Question 4	0.981	0.930	0.980	0.915	0.983	0.962
Goods, excluding motor vehicles						
Question 1	1.025	0.946	1.033	0.987	1.016	0.900

Source: Authors' calculations.

Notes: The table reports the ratio of the root-mean-squared forecasting error for the equation containing the question to the root-mean-squared forecasting error for the equation without the question; a number less than one indicates that including the question improves the forecast accuracy relative to the baseline model for that particular category of consumption. Out-of-sample evaluation periods appear at the top of each column; the initial estimation period begins with the first quarter of 1968 and ends with the fourth quarter of 1981.

ability. Including this question in the forecasting equation consistently improves the out-of-sample forecasts of total personal consumption expenditure during this period. It also improves the model's out-of-sample performance in both subperiods—most notably in the 1980s. The out-of-sample forecasting power of the Conference Board's question four corroborates the in-sample finding that questions about job availability typically have the most predictive power.

Other results show that the Conference Board's questions one through four generally improve forecasts in every period for motor vehicle expenditure. Michigan's questions one, two, and four are also useful for forecasting motor vehicle spending.

To summarize, like the in-sample tests, the out-of-sample results show that some survey questions have more predictive power than others. Questions that ask about consumers' perceptions of job availability typically have the most explanatory power for future movements in consumption, whereas questions that ask about buying conditions or financial conditions today relative to the past appear to have much less explanatory power.

Conclusion

This paper investigates the impact of consumer attitudes on consumer spending. The purpose of our empirical analysis is to compare the forecasting power of two widely followed measures of consumer perspectives—the Conference Board Consumer Confidence Index and the University of Michigan Index of Consumer Sentiment. We also discuss the ways in which the surveys underlying these measures differ and test whether certain types of survey questions are particularly important for predicting consumer expenditures.

We find that lagged values of the Conference Board Consumer Confidence Index provide information about the future path of spending that is not captured by lagged values of the Michigan Index of Consumer Sentiment, labor income, stock prices, interest rates, or the spending category itself. These results contrast with those of other researchers, such as Carroll, Fuhrer, and Wilcox (1994), who find that consumer attitudes, as measured by the University of Michigan index, contribute little additional information.

The most obvious implication of our empirical results is that forecasts of total personal consumer spending may be made more accurate by utilizing the Conference Board's Consumer Confidence Index. Forecasts are often improved either by replacing the Michigan index with the attitudinal indicator from the Conference Board or by combining the Conference Board data with more conventional economic variables such as income, consumption, and financial indicators.

We also find that the general superiority of the Conference Board index for forecasting consumption appears to be related to the types of questions that make up the survey. The two Conference Board questions that ask specifically about job prospects in the respondent's area exhibit the most predictive power. By contrast, in the Michigan index, the two questions that focus on current buying conditions or financial conditions in the recent past display little incremental forecasting power. Thus, when the surveys of consumer attitudes reveal a major shift in sentiment, policymakers and forecasters might wish to pay close attention to the questions that generated this response. For example, a surge in consumer confidence that is largely driven by the questions about future job availability might suggest greater potential for increased consumer spending than a surge in confidence that is driven by other questions. Consumers seem to spend more when they feel good about future job prospects than they do when they think business conditions are favorable.

We have left at least one important topic for future research: the issue of what theoretical model might account for the spending-confidence correlations we have found. We caution that our results do not prove that consumer attitudes cause changes in consumer spending. Although our analysis explicitly controls for economic fundamentals regarded as important determinants of aggregate consumption growth, the possibility remains that some other variable may be driving the confidence-spending correlations found here. Nevertheless, our results suggest that consumer confidence can help predict consumption, and that consumer attitudes may also act as a catalyst for economic fluctuations.

Appendix A: Correlation Matrix

	University of Michigan Index			Conference Board Index		
	Total	Expectations	Present Conditions	Total	Expectations	Present Conditions
Michigan total	1.00	0.96	0.90	0.69	0.71	0.48
Michigan expectations		1.00	0.75	0.68	0.80	0.42
Michigan present conditions			1.00	0.59	0.45	0.51
Conference Board total				1.00	0.71	0.91
Conference Board expectations					1.00	0.34
Conference Board present conditions						1.00

Appendix B: Definitions of the Estimation Variables

Consumption

We examine five categories of real personal consumption expenditure: total expenditure; motor vehicles; goods, excluding motor vehicles; services; and durables, excluding motor vehicles. The quarterly data are from the U.S. Department of Commerce, Bureau of Economic Analysis.

Labor Income

Labor income is defined as wages and salaries plus transfers minus personal contributions for social insurance. These quarterly components are from the Department of Commerce's National Income and Product Accounts.

Interest Rates

The interest rate is the three-month Treasury bill rate, reported monthly by the Board of Governors of the Federal Reserve System. The data are quarterly averages.

Stock Prices

Stock prices equal the Standard and Poor's 500 composite stock price index (1941-43=10). The data are quarterly averages.

Price Deflator

Nominal labor income and the Standard and Poor's 500 index are deflated by the personal consumption expenditure implicit price deflator (1992=100). The data are reported quarterly in the National Income and Product Accounts. The data reflect revisions in September 1993.

ENDNOTES

The authors are grateful to Jeffrey Fuhrer, Jonathan McCarthy, Patricia Mosser, Gabriel Perez Quiros, Robert Rich, Rae Rosen, Christopher Sims, Charles Steindel, and Egon Zakrajšek for helpful comments, and to Beethika Khan for excellent research assistance.

1. Of course, there may have been other proximate causes of the 1990-91 recession such as the Persian Gulf War and commodity-price or bank-loan supply shocks associated with the war.

2. The Michigan index begins with quarterly data in 1952; the Conference Board index with bimonthly data in 1967.

3. Early investigators of the explanatory power of consumer confidence include Fair (1971), who links the University of Michigan index with both durable and nondurable consumer expenditures, and Mishkin (1978), who argues that the Michigan index may be a good proxy for the consumer's subjective assessment of the probability of future financial distress. More recent work analyzing the Michigan index can be found in Carroll and Dunn (1997), Carroll, Fuhrer, and Wilcox (1994), Fuhrer (1993), Leeper (1992), and Matsusaka and Sbordone (1995).

4. We leave for future research the question whether some theoretical model might explain the predictive power of consumer attitudes for consumption.

5. Because the Conference Board index includes a question about nominal income, it may overstate "confidence" during periods of high inflation.

6. This difference in time horizons may have some effect on response patterns and hence on index results.

7. There may be some sample selection bias in both surveys, but any such bias is assumed to be constant over time and so has virtually no effect on the indexes.

8. Because of differences in index construction, discussed earlier, the Conference Board's index has a wider range of movement than Michigan's. However, on a standardized basis, the Conference Board's index is significantly *less* volatile—that is, it has a higher signal-to-noise ratio than Michigan's index.

9. As noted earlier, the University of Michigan quarterly data are available from 1952, while the Conference Board data do not begin until the first quarter of 1967. To maintain a basis of comparison across regressions, we use the largest possible sample for which both indexes are available.

10. Estrella and Hardouvelis (1989) have established the forecasting power of this "term structure" spread for several real variables.

11. The growth in spending on durable goods may be positively autocorrelated, with the error term following a first-order moving-average process (see Mankiw [1982]). First-order autocorrelation in the error term may cause the error term to be correlated with the one-period-lagged endogenous variable, a condition that could skew in-sample statistical tests of the joint marginal significance of the explanatory variables (the reported *p*-values). To address this problem, we explicitly model the error term, ε_t, following an MA(1) process in the in-sample regressions. This strategy is derived from Carroll, Fuhrer, and Wilcox (1994). Allowing for an MA(1) in the error term requires nonlinear estimation, and we use nonlinear least squares in our in-sample estimation of equation 1 and in the confidence-augmented equations that follow. The coefficient on the lagged-moving-average term generally has the expected negative sign. For example, for total real personal consumption expenditures and the confidence-augmented equation using the Michigan index, the coefficient is estimated at -0.8, with a standard error of 0.1. See Carroll, Fuhrer, and Wilcox (1994).

12. We do not report results for the present conditions component because preliminary tests indicated that the expectations component of both indexes typically exhibited greater forecasting power.

13. Previous research (for example, Leeper [1992]) suggests that consumer confidence may be linked to economic indicators such as unemployment and industrial production largely because of unusually volatile movements in consumer attitudes during the Persian Gulf War and the 1990-91 recession. To control for this possibility, we include a dummy variable set equal to one in the quarters corresponding to the 1990-91 recession. We then eliminate the dummy variable and perform out-of-sample forecasts over several evaluation periods using the beginning of the 1990-91 recession as a break.

14. Adding a dummy variable for the third quarter of 1980 to account for credit controls does not significantly alter the results; in the sample controlling for the 1990-91 recession, the incremental adjusted R^2 is .09, and the lags of the Conference Board index are jointly significant at better than the 1 percent level. The adjusted R^2 from a regression of total personal consumption expenditure growth on the controls alone is approximately .40.

15. By regressing consumption growth on four lags of the overall index, we implicitly restrict the coefficient on each component (relative to its share in the overall index) to be the same. One question to consider is whether the forecasting power of the Conference Board index would be improved by regressing the consumption growth category on the

Note 15 continued

expectations and present conditions components separately. We investigated this question but found that the incremental adjusted R^2 increased significantly in just one category: motor vehicle spending. In that equation, when we added four lags of each component separately, the increment to the adjusted R^2 increased to 16 percent, from 5 percent.

16. Unlike the overall indexes, however, the question-level indexes are not pegged to a base year. Question-level data for the University of Michigan survey come from the Board of Governors of the Federal Reserve System. We thank Lynn Franco of the Conference Board for providing us with data on the Conference Board questions.

17. Conference Board question one also has statistically significant explanatory power in the motor vehicle expenditure equation. However, the increment to the adjusted R^2 is quite modest and considerably smaller than that produced by the other questions for this expenditure category. Michigan's question one is a statistically significant predictor of motor vehicle spending at the 10 percent level but not at the 5 percent level.

18. This finding makes sense because motor vehicles are more likely to be financed using long-term credit than are other durable goods.

19. Note that the first subperiod does not include the 1990-91 recession, so that the recession cannot explain the predictive power of the Conference Board index.

20. There are other problems with statistical tests for equal forecast accuracy. Harvey, Leybourne, and Newbold (1997) have documented the severe size problems of the standard Diebold-Mariano test. Their modified test, used in this study, goes part of the way toward fixing the size problems but does not eliminate them.

REFERENCES

Bram, Jason, and Sydney Ludvigson. 1997. "Does Consumer Confidence Forecast Household Expenditure? A Sentiment Index Horse Race." Federal Reserve Bank of New York Research Paper no. 9708.

Campbell, John Y., and Gregory Mankiw. 1989. "Consumption, Income, and Interest Rates: Reinterpreting the Time Series Evidence." In O. Blanchard and S. Fischer, eds., NBER MACROECONOMICS ANNUAL. Cambridge: MIT Press.

Carroll, Christopher D., and Wendy Dunn. 1997. "Unemployment Expectations, Jumping (S,s) Triggers, and Household Balance Sheets." NBER MACROECONOMICS ANNUAL. Cambridge: MIT Press.

Carroll, Christopher D., Jeffrey C. Fuhrer, and David W. Wilcox. 1994. "Does Consumer Sentiment Forecast Household Spending? If So, Why?" AMERICAN ECONOMIC REVIEW 84, no. 5: 1397-1408.

Clark, Todd E. 1996. "Finite-Sample Properties of Tests for Forecast Equivalence." Federal Reserve Bank of Kansas City Research Paper no. 96-03.

Estrella, Arturo, and Gikas A. Hardouvelis. 1989. "The Term Structure as a Predictor of Real Economic Activity." Federal Reserve Bank of New York Research Paper no. 8907.

Fair, R. C. 1971. "Consumer Sentiment, the Stock Market and Consumption Functions." Princeton University Econometric Research Program Research Memorandum 119.

Flavin, Marjorie. 1981. "The Adjustment of Consumption to Changing Expectations about Future Income." JOURNAL OF POLITICAL ECONOMY 89, no. 5: 974-1009.

Fuhrer, Jeffrey C. 1993. "What Role Does Consumer Sentiment Play in the U.S. Economy?" Federal Reserve Bank of Boston NEW ENGLAND ECONOMIC REVIEW, January-February: 32-44.

Harvey, David, Stephen Leybourne, and Paul Newbold. 1997. "Testing the Equality of Prediction Mean Squared Errors." INTERNATIONAL JOURNAL OF FORECASTING 13, no. 2 (June): 281-91.

Leeper, Eric M. 1992. "Consumer Attitudes: King for a Day." Federal Reserve Bank of Atlanta ECONOMIC REVIEW 77, no. 3 (July-August): 1-16.

Mankiw, Gregory. 1982. "Hall's Consumption Hypothesis and Durable Goods." JOURNAL OF MONETARY ECONOMICS 10, no. 3: 417-25.

Matsusaka, John G., and Argia M. Sbordone. 1995. "Consumer Confidence and Economic Fluctuations." ECONOMIC INQUIRY 33, no. 2 (April): 296-318.

Mishkin, Frederic S. 1978. "Consumer Sentiment and Spending on Durable Goods." BROOKINGS PAPERS ON ECONOMIC ACTIVITY, no. 1: 217-32.

Wilcox, David W. 1992. "The Construction of U.S. Consumption Data: Some Facts and Their Implications for Empirical Work." AMERICAN ECONOMIC REVIEW 82, no. 4: 922-41.

Vertical Specialization and the Changing Nature of World Trade

David Hummels, Dana Rapoport, and Kei-Mu Yi

The world's economies have become increasingly integrated and increasingly global. Among the most important and often cited features of the rise in globalization is the enormous growth in the export and import shares of GDP since World War II. In the United States, international trade—that is, exports plus imports—accounted for 23.9 percent of GDP in 1996, up from 9.2 percent in 1962.[1] Worldwide, the merchandise export share of production has more than doubled over the last forty-five years, while the manufactured export share of production has almost quadrupled (Chart 1). Most countries—emerging nations as well as highly developed economies—have experienced increases in their export share of GDP (Chart 2). Clearly, a greater number of countries are trading more today than in the past.

David Hummels is an assistant professor of economics at the University of Chicago's Graduate School of Business; Dana Rapoport is an assistant economist and Kei-Mu Yi an economist at the Federal Reserve Bank of New York.

Another significant feature of increased globalization is the internationalization of production. Rather than concentrate production in a single country, the modern multinational firm uses production plants—operated either as subsidiaries or through arm's-length relationships—in several countries. By doing so, firms can exploit powerful locational advantages, such as proximity to markets and access to relatively inexpensive labor. There are currently more than 39,000 parent firms and 279,000 foreign affiliates worldwide, with a total foreign direct investment (FDI) stock equal to $2.7 trillion in 1995, compared with $1.0 trillion in 1987. Moreover, the value added of foreign affiliates—that is, their sales less materials costs—accounted for 6 percent of world GDP in 1991, a 300 percent increase from 1982 (United Nations Conference on Trade and Development 1996).

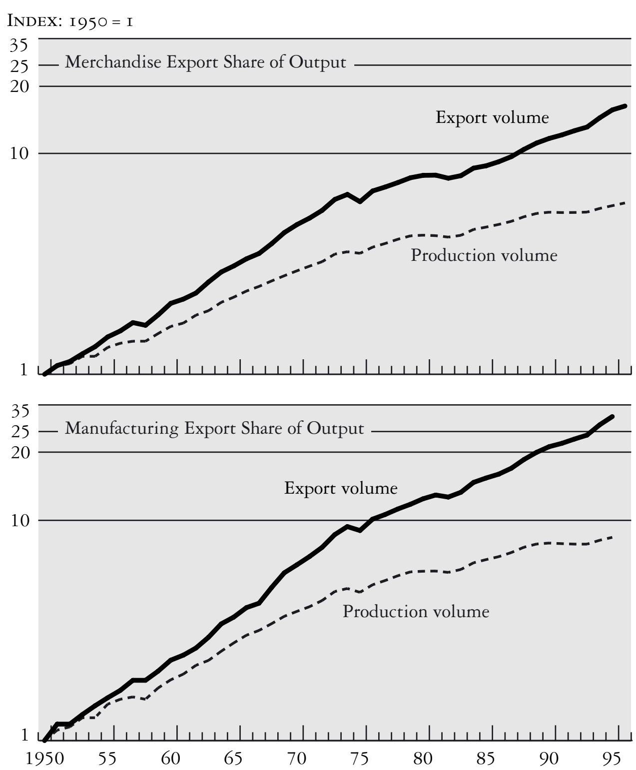
Increased international production, however, does not always lead to increased international trade. For instance, if a country's firms serve markets abroad through production facilities in each country—rather than through exports from the home country—trade may actually decrease as international production rises. International production *will* be associated with increased trade when countries are vertically linked—that is, when international production prompts countries to specialize in particular stages of a good's production. In that case, a sequential mode of production arises in which a country imports a good from another country, uses that good as an input in the production of its own good, and then exports its good to the next country; the sequence ends when the final good reaches its final destination. We use the term "vertical specialization" to describe this mode of production.[2] By comparison, in a horizontal-specialization scenario, countries trade goods that are produced from start to finish in just one country.

In this article, we shed light on the globalization of international production and trade by demonstrating the increasingly important role vertical specialization plays in international trade. We use case studies and input-output tables to calculate the level and growth of vertical-specialization-based trade, which we define as the amount of imported inputs embodied in goods that are exported. The case studies—the United States–Canada Auto Agreement of 1965, Mexico's maquiladora trade with the United States, electronics trade between Japan and Asia, and trade involving Opel's subsidiary in Spain—allow us to quantify the amount of vertical-specialization-based trade.[3] In all of the case studies, our findings indicate that vertical specialization has increased sharply in recent years: in the Japan-Asia electronics trade, for example, it increased 900 percent between 1986 and 1995.

To show that the results of our case studies can be generalized, we use input-output tables to calculate estimates of vertical-specialization-based trade in ten developed countries from the Organization for Economic Cooperation and Development (OECD). We find that by the beginning of the present decade, 14.5 percent of all trade in these countries was vertical-specialization-based—a 20 percent increase

Chart 1

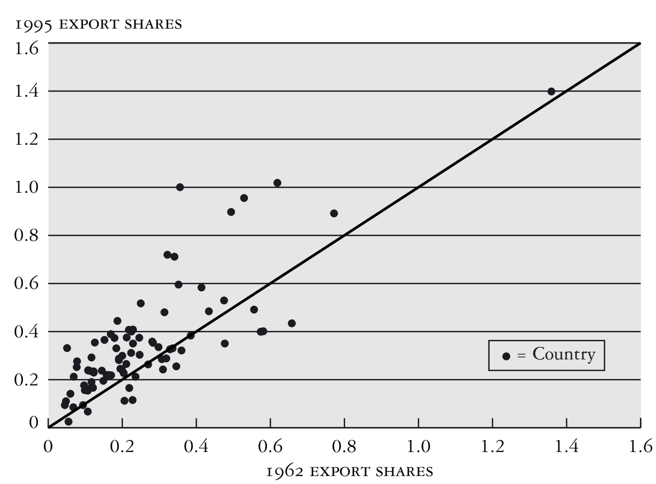
World Production and Export Volumes



Source: World Trade Organization Secretariat.

Chart 2

Countries' 1962 and 1995 Export Shares of GDP: Most Countries' Export Shares Increased



Source: Authors' calculations, based on data from International Monetary Fund, *International Financial Statistics Yearbook*.

Notes: Each dot represents a different country. If a country lies above and to the left of the 45° line, then its 1995 export share is larger than its 1962 export share. It is clear that this is the case for a majority of the countries. Furthermore, this phenomenon is true for all types of countries: Countries as distinct as Bangladesh, the Congo, Germany, Ireland, Korea, Malaysia, and the United States all lie above the 45° line.

from the late 1960s and early 1970s. Thus, while the majority of trade continues to be horizontal, vertical-specialization-based trade is making significant gains.

Analysis of our OECD data reveals a strong statistical association between the increased vertical specialization share of total trade and the rising trade shares of GDP. In addition, it shows that the industries accounting for most of the increase in the vertical specialization share of total trade—chemicals, and machinery and equipment—also account for most of the increase in overall trade as a share of GDP. Increases in vertical-specialization-based trade are also found to account for more than 25 percent of the increase in total trade in most of our ten OECD countries.

Vertical specialization occurs when a country uses imported intermediate parts to produce goods it later exports. This definition captures the idea that countries link sequentially to produce a final good.

Our study also considers some implications of the increase in vertical specialization for trade policy. Although a detailed examination of policy issues is beyond the scope of this article, our results lead to two tentative conclusions. First, even though tariff and nontariff barriers worldwide are now quite low, especially among the developed countries, vertical specialization can magnify the gains that are achieved by lowering these barriers even further. Second, vertical specialization has helped make the linkages between foreign direct investment policy and trade policy stronger than ever. The trade gains from vertical specialization can therefore be realized when countries place greater emphasis on eliminating FDI restrictions.

In the next section, we define vertical specialization more precisely and relate it to other important production concepts, such as outsourcing, vertical foreign direct investment, and vertical integration. We then present our case study and input-output table evidence of the increased importance of vertical specialization. We conclude with a discussion of the possible causes of vertical specialization, as well as its trade policy implications.

Vertical Specialization

The story of globalization is a story about specialization. Today, countries focus more and more on producing a relatively narrow range of goods and services. They exchange the fruits of their specialization for other goods and services. The traditional notion of specialization is horizontal—firms or countries become adept at producing particular goods and services from scratch and then export them. We show, however, that an increasingly significant characteristic of world trade is vertical specialization.

Three conditions must hold for our definition of vertical specialization to occur: (1) a good must be produced in multiple sequential stages, (2) two or more countries must specialize in producing some, but not all, stages, and (3) at least one stage must cross an international border more than once.[4] In other words, vertical specialization occurs when a country uses imported intermediate parts to produce goods it later exports. This definition captures the idea that countries link sequentially to produce a final good.

To obtain a quantitative measure of the amount of trade due to vertical specialization, we define vertical-specialization-based trade to be the value of imported intermediates embodied in a country's exports, multiplied by two. We multiply by two because imported intermediates are counted twice: once as imports and once as embodied in exports. We calculate this trade as:

$$(1) \quad (\textit{fraction of gross production that is imported intermediates}) \times (\textit{exports}) \times (2)$$

or, equivalently, as:

$$(1a) \quad (\textit{imported intermediates}) \times (\textit{fraction of gross production that is exported}) \times (2).$$

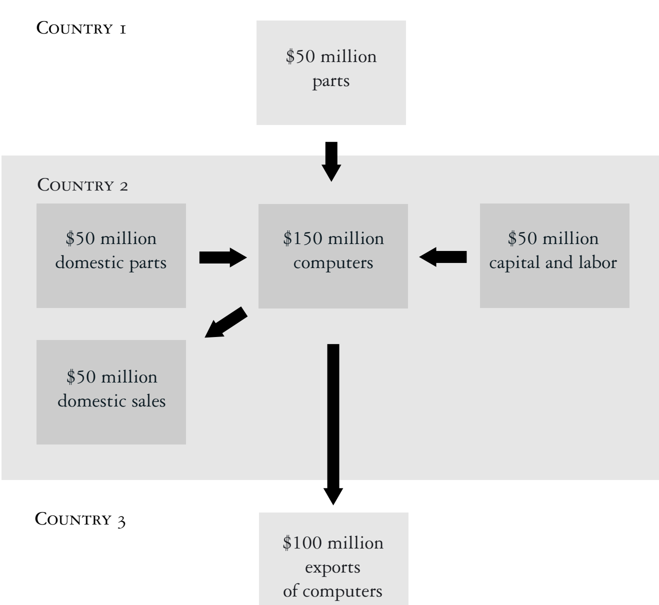
From the above formulas, we can see that as the fraction of gross production that is imported intermediates and/or the fraction of gross production that is exported increases, the fraction of trade that is vertical-specialization-based will increase.

The exhibit below offers a good example. Country 2 imports $50 million of parts from Country 1 and, after producing computers, exports $100 million of the computers to Country 3. Applying equation 1, we see that vertical-specialization-based trade for Country 2 is $(\$50/\$150) \times \$100 \times 2 = \$200/3$ million, which is twice the value of imported intermediates embodied in exports. Because Country 2’s total trade is $150 million, vertical-specialization-based trade thus accounts for 44 percent of its total trade. However, if one or both of the imported intermediates and exports were zero, vertical-specialization-based trade would also be zero.

Vertical-specialization-based trade is clearly related to trade in intermediate goods, which has also risen sharply in recent decades. However, our definition makes clear that vertical-specialization-based trade can include trade in final goods, as long as some imported intermediates are used to produce those goods. The above example also shows that imports of intermediate goods would not count as vertical-specialization-based trade if the good produced with the imported intermediates was not exported.

Vertical specialization is related to several production concepts including outsourcing, vertical integration, and vertical FDI, all of which have garnered much attention in academic research and the popular press. Outsourcing is the relocation of one or more stages of a good that was formerly produced entirely in the home country (see box). Vertical integration and vertical FDI are activities in which multinational firms locate different stages of production of a good or goods in different countries. These concepts are similar to vertical specialization because they are all concerned with the location of production. The main distinction, however, is that vertical specialization concerns the activities of countries, while outsourcing, vertical integration, and vertical FDI involve the behavior of multinational firms. While firm-level production activities represent ways in which country-level vertical specialization can occur, outsourcing and vertical integration and FDI need not imply vertical specialization, and vice versa.[5] We later show that the industries in which multinational firms are engaged—manufacturing industries, especially

An Example of Vertical-Specialization-Based Trade



Outsourcing and Vertical Specialization

We define outsourcing as the relocation of one or more stages of the production of a good from the home country. Labor, capital, and/or technology can be transferred in the process. While we regard outsourcing as the act of relocation, others have viewed it more generally as reliance on imported inputs. The examples below illustrate the distinction between vertical specialization and both types of outsourcing:

1. Suppose a firm relocates production of computer components to another country and imports these components from that country. The firm then completes the production of the computers but does not export them. In this case, outsourcing—as we define it—has occurred, but vertical specialization has not. However, if the country does export the final goods, both outsourcing and vertical specialization have occurred.
2. Suppose a country manufactures computers and some of the intermediate inputs are imported. In this case, outsourcing—as others define it—has occurred. If no computers are exported, there is no vertical specialization; if computers are exported, vertical specialization has occurred.
3. According to our definition, outsourcing is not present in the second example. Either way, however, vertical specialization and outsourcing are distinct concepts.

chemicals, machinery, and equipment—are those in which the share of total trade that is vertical-specialization-based is the largest.[6]

Empirical Evidence: Four Case Studies

To construct empirical measures of vertical-specialization-based trade, ideally we would use data on the production process and direction of trade flow for every stage of each good traded in the world economy.[7] Unfortunately, these data are impossible to obtain. We can, however, construct detailed estimates of vertical trade on a case-by-case basis. (Appendix A provides additional details on the data sources for our four case studies.)

Implicit in all of our case study calculations is the assumption that countries divide production into two stages—intermediate goods production and final goods production—with one stage occurring in each country.[8] To the extent that countries divide production into more than two stages, our calculations underestimate the amount of vertical trade. For example, imagine that the United States used pistons imported from Canada to produce engines that are then exported to Canada, where they are assembled into final motor vehicles that are exported back to the United States. In this instance, our calculations would miss one set of trade flows, or one "border crossing."

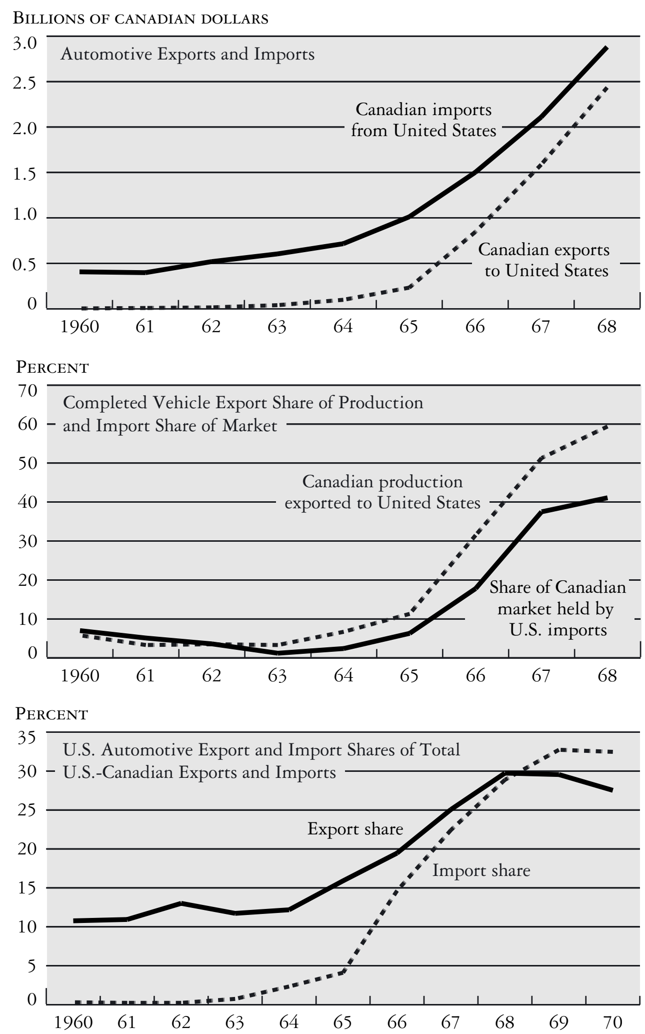
Our first two case studies illustrate bilateral relationships—that is, relationships in which one country exports goods to a second country, which uses them as inputs to produce goods that are exported back to the first country. In the other two studies, the second country, rather than exporting the goods back to the first country, exports them to a third country.

The 1965 United States–Canada Auto Agreement

Before the 1965 United States–Canada Auto Agreement, auto trade between the two countries was virtually nonexistent. Tariffs were significant: 17.5 percent on Canadian automotive imports from the United States and 6.5 to 8.5 percent on U.S. automotive imports from Canada. Canadian auto manufacturers (affiliates of GM, Ford, Chrysler, and American Motors Corporation) produced exclusively for the Canadian market, and almost all vehicles sold in Canada were also made there. The 1965 agreement reduced the tariffs facing producers to zero.[9] Viewing the United States and Canada as one integrated market after the agreement, U.S. auto companies immediately consolidated production. In Canada, production was narrowed to just a few models, with the output serving the entire North American market. Just four years after the agreement, auto trade soared (Chart 3). The share of Canadian vehicles exported to

Chart 3

U.S.-Canadian Auto Trade: Before and After the 1965 Auto Agreement



Sources: Top panel: Beigie (1970, p. 71, Table 13); middle panel: Beigie (1970, p. 72, Table 14); bottom panel: U.S. Department of Commerce, Bureau of Economic Analysis, *Survey of Current Business*, International Monetary Fund, *Direction of Trade*.

the United States leaped from 7 percent to 60 percent, and the share of the Canadian automobile market consisting of imported cars jumped from 3 percent to 40 percent (Beigie 1970, pp. 4-5). The bottom panel of the chart shows that the automobile share of total U.S.-Canadian trade rose immediately, from approximately 8 percent to 30 percent.[10]

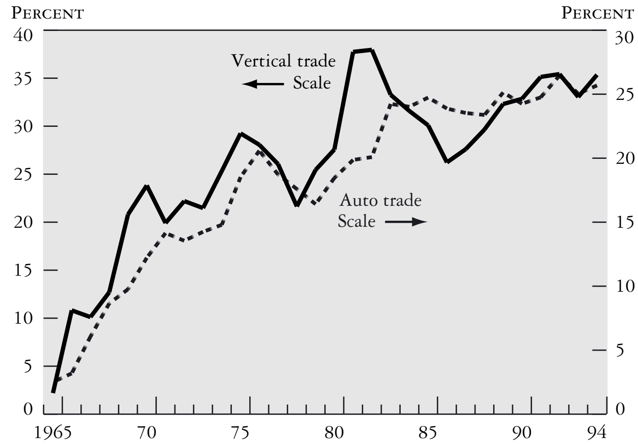
These events seem like a textbook example of traditional horizontal specialization, in which there is just more trade in motor vehicles.[11] Nevertheless, the basic data provide a hint that vertical specialization also occurred. Sixty percent of U.S. auto exports to Canada are engines and parts, while 75 percent of U.S. auto imports from Canada are finished cars and trucks (U.S. Department of Commerce 1994-96). To proceed further, we estimate the level of U.S.-Canadian vertical trade following the auto agreement using data from *Ward's Automotive Yearbook*, the United Nations Comtrade database, and the U.S. Department of Commerce's Bureau of Economic Analysis (BEA). The United Nations and BEA trade data separate parts trade from vehicles trade, a distinction that is key to our calculation. Our calculation has two steps. First, we attribute trade in auto parts in 1964, before the agreement, entirely to the auto repair market, for which there is no vertical specialization. We also assume that the ratio of repair market trade to total U.S. auto and truck output is constant at its 1964 level in order to calculate estimates of repair trade in future years. We subtract this amount of trade from the overall trade figures to obtain an estimate of parts trade owing specifically to the agreement (Appendix A). Second, we determine the amount of vertical trade generated by the auto agreement, which equals:

(2) $$2 \times \{ [(\textit{adjusted}\,)\ \textit{Canadian parts imports}] \times [\textit{fraction of Canadian vehicle production exported to United States}] + [(\textit{adjusted}\,)\ \textit{U.S. parts imports}] \times [\textit{fraction of U.S. vehicle production exported to Canada}] \}.$$

Note that we calculate vertical trade in both directions—that is, from the United States to Canada to the United States, and from Canada to the United States to Canada. The fraction of Canadian vehicle production exported to the United States is currently about 80 to 90 percent; by contrast, only a small fraction of U.S. vehicle production is exported to Canada. This means that the vast majority of vertical trade consists of the U.S.-Canadian-U.S. flow. Chart 4 shows the percentage of total automotive trade from 1965 to 1994 that is vertical trade generated by the auto agreement. By 1971, vertical trade had risen from 0 percent to more than 20 percent of total auto trade, and it has continued to trend upward. In recent years, vertical trade has accounted for more than 35 percent of U.S.-Canadian auto trade, or about $30 billion. Chart 4 also shows that the share of vertical trade in total U.S.-Canadian auto trade is highly correlated with total U.S.-Canadian auto trade as a fraction of U.S. auto output; the correlation coefficient is 0.82.

Chart 4

VERTICAL TRADE AS A PERCENTAGE OF AUTO TRADE AND AUTO TRADE AS A PERCENTAGE OF AUTO OUTPUT: THE UNITED STATES AND CANADA



Source: Authors' calculations, based on data from the United Nations Statistical Division, the U.S. Commerce Department's Bureau of Economic Analysis, and *Ward's Automotive Yearbook*.

Note: The data are described in greater detail in Appendix 1.

MEXICO'S MAQUILADORAS

Mexico's maquiladoras are non-Mexican-owned production plants that complete processing or secondary assembly of imported components for export.[12] These plants benefit from Mexican laws that exempt from Mexican tariffs parts and materials imported by Mexico for use in maquiladoras.

Also, U.S. firms that use maquiladoras receive favorable tax treatment from the United States. Under U.S. law, the U.S. components of maquiladora-made goods exported back to the United States are exempt from U.S. tariffs.[13] Consequently, the only part of the two-way transaction that is dutiable is the Mexican value added in the goods exported back to the United States. The net effect of these policies is that U.S. firms increasingly have turned to vertical specialization by outsourcing to the maquiladoras a large fraction of manufactured goods assembly.

The principal maquiladora industries are electric/electronics, transportation equipment, and textiles, which together employ more than 73 percent of all maquiladora workers and account for 81 percent of total maquiladora production. The electric/electronics industry is the largest, accounting for almost half of total maquiladora production in 1994. The transportation sector has grown the fastest in recent years, increasing its share of employment from 10 percent in 1982 to 22 percent in 1995.

From the maquiladoras' inception in 1965 until the early 1980s, their growth was steady but not striking. However, propelled by the greater importance given to them by Mexico's de la Madrid administration, maquiladoras grew considerably starting in the mid-1980s. From 1985 to 1997, employment growth in maquiladoras averaged 12.6 percent per year, and almost 900,000 workers were employed in 1997. The maquiladoras' increases in gross production were equally striking, averaging an annual growth rate of 19.7 percent during the same period; gross production was $44 billion in 1997.[14] The growth in production has been accompanied by strong growth in total bilateral trade as a share of Mexico's GDP (Chart 5). Since the late 1980s, U.S. maquiladora imports have represented 45 percent of total U.S. imports, and 60 percent of total non-oil U.S. imports, from Mexico (Hufbauer and Schott 1992, pp. 96-7).

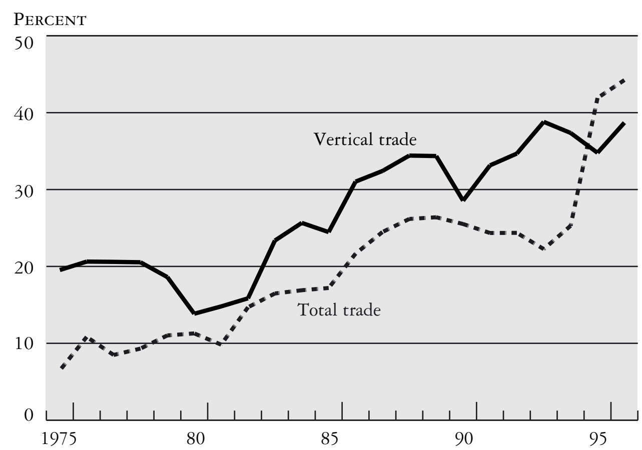
Our maquiladora data include imported intermediates and gross production. In addition, we know that almost all imported intermediates are from the United States and almost all production is exported there; hence, we assume that these shares are 100 percent.[15] We compute vertical-specialization-based trade only for the flow of Mexican imported intermediates from the United States used to produce goods exported back to the United States because we do not have data on U.S. imported intermediates from Mexico that are used to produce goods exported back to Mexico (vertical trade flows in the latter direction are probably not large anyway).

Our calculations indicate that vertical trade has increased significantly. Between 1975 and 1979, the share of total U.S.-Mexican trade attributable to maquiladora vertical trade averaged about 20 percent per year (Chart 5). This share rose to an average of 25 percent in the following decade and of 35 percent in the first half of the 1990s, reaching 39 percent in 1996. Such trade in 1996 represented about $57 billion. Because there is surely vertical trade originating from nonmaquiladora channels, we contend that at least half of U.S.-Mexican trade could be due to vertical specialization.

Our analysis also suggests a relationship between the growth in vertical trade and the increase in total bilateral trade as a share of Mexico's GDP (Chart 5). Vertical trade and total bilateral trade have followed similar,

Chart 5

VERTICAL TRADE AS A PERCENTAGE OF TOTAL TRADE AND TOTAL TRADE AS A PERCENTAGE OF GDP: THE UNITED STATES AND MEXICO



Sources: Authors' calculations, based on data from Instituto de Estadistica, Geografia e Informatica, International Monetary Fund, *International Financial Statistics* CD-ROM and *Direction of Trade Statistics* CD-ROM, Banco de Mexico, *The Mexican Economy*.

although not identical, paths. The correlation coefficient between the two variables is significantly positive, 0.83.

Japan-Asia Electronics Trade

To reduce costs, many of Japan's manufacturing industries have been rapidly outsourcing different stages of production, especially final assembly, to Southeast Asia and other countries. In 1996, almost 70 percent of Japanese offshore electronics production facilities were located in just nine developing Asian countries. As of 1995, offshore workers accounted for almost 40 percent of total Japanese electronics industry employees, up from just 25 percent in 1989. It is no surprise, then, that offshore production has surpassed domestic production of both color televisions (in 1988) and VCRs (in 1994).

Using data from the Electronic Industries Association of Japan and the Japan Electronics Bureau, we show patterns of production and exports for the Japanese electronics industry between 1985 and 1995 (Chart 6). We see in the top panel that the export share of components and devices has increased, while the export share of consumer and industrial equipment has remained virtually constant or even decreased during this period. Developing countries in Asia are playing a greater role in the rising importance of components. As of 1995, exports of components to Asia accounted for more than three-fourths of all exports there, more than one-half of all exports of components, and more than one-third of total electronics exports. These components are used primarily for production of other components or final goods such as VCRs and color televisions. Most of this offshore production is then exported back to Japan or to third countries such as the United States.

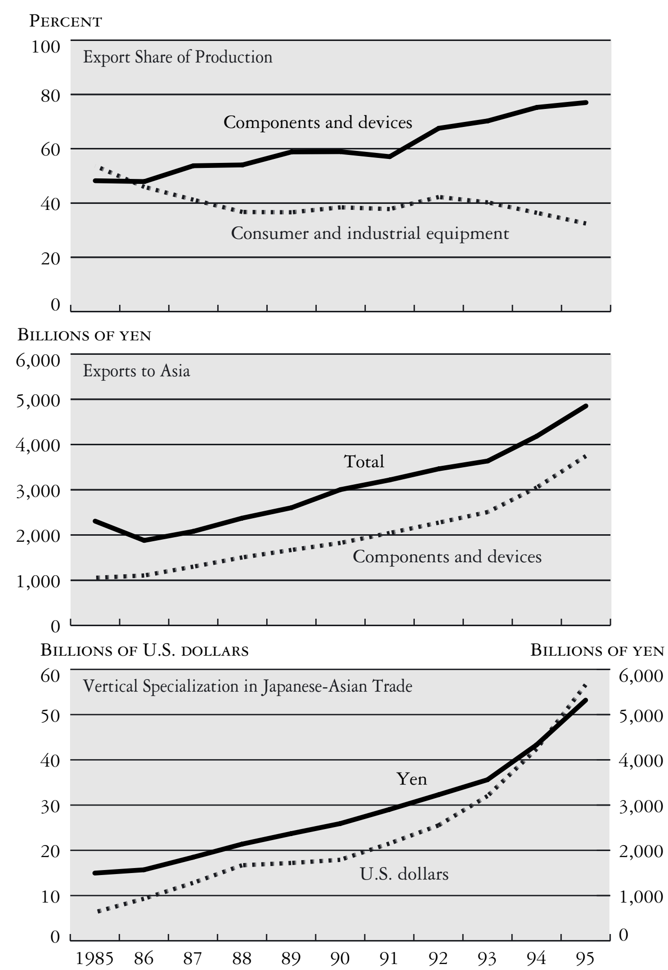
We make two assumptions to estimate the amount of electronics vertical trade for the countries in which the Japanese electronics industry relocated its production. First, we assume that all electronic components imported from Japan are used as inputs for further production. Second, using Wells' (1993) finding that Japanese electronics subsidiaries in Indonesia export 71 percent of their production, we assume that this percentage applies to all Asian countries with Japanese subsidiaries. Under these assumptions, vertical trade equals

$$(3) \qquad 2 \times [\textit{imports of components from Japan}] \times [0.71].$$

Applying this formula, we find that in the last ten years vertical-specialization-based trade has almost quadrupled in yen terms and has increased ninefold in dollar terms; as of 1995, it was approximately $55 billion (Chart 6, bottom panel). By contrast, total electronics exports from Japan during this period increased by only 23 percent in yen terms and by about 81 percent in dollar terms.

Chart 6

Japanese Electronics Industry



Sources: Electronic Industries Association of Japan, *Facts and Figures on the Japanese Electronics Industry* and *Perspectives on the Japanese Electronics Industry*, Japan Electronics Bureau, International Monetary Fund, *International Financial Statistics Yearbook*.

Opel's Subsidiary in Spain

Opel, General Motors' affiliate in continental Europe, began operations in Spain in 1982. As of 1994, Opel España made about 22 percent of Spain's total production of 1.8 million passenger cars. From the beginning, Opel España was an important participant in vertical trade, relying heavily on imported inputs to produce automobile parts and final vehicles, most of which were exported. To calculate vertical trade, we use 1983-95 Opel España data on net sales of vehicles and parts, exports of vehicles and parts, and imported parts. As in the two previous case studies, we can calculate this trade only in one direction. (Because countries in the European Union are likely to have extensive production and trade networks, vertical trade in the other direction—that is, Spain exporting intermediate goods to other countries and then importing goods embodying those intermediates—may be significant.) With the data we do have, we estimate the amount of Opel España's vertical-specialization-based trade to be both significant and increasing: $0.6 billion in 1983, $1.8 billion in 1993, $2.7 billion in 1994, and $3.6 billion in 1995.

Using additional data on Spain's auto exports, we can estimate the contribution of vertical trade to the country's overall auto trade. We know that Spain's other auto companies—affiliates of American or European corporations—all export a somewhat smaller fraction of their passenger car production than Opel España, about 70 percent as opposed to 90 percent. Using Opel España's market share of 22 percent, and assuming that these other companies rely on imported inputs to the same degree as Opel, we estimate Spain's total vertical trade in autos to be $13.5 billion in 1995, up from $6.8 billion in 1993 and $10.1 billion in 1994. When we compare these latter figures with total Spanish auto trade of $21 billion in 1993 and $25 billion in 1994, we estimate that at least 40 percent of Spanish auto trade is vertical trade.[16]

Empirical Evidence: Input-Output Tables

We have established the quantitative significance of vertical specialization for our case studies. Next, we ask whether the conclusions reached through case studies of specific operations and products can be extended to entire countries. To resolve this issue, we use input-output tables to calculate vertical-specialization-based trade. Input-output tables characterize, in matrix fashion, the interrelationships among industries in a country's economy. The tables report, for example, how much of the steel industry's output is used as an intermediate input in the motor vehicles industry and vice versa. They also report the gross output and value added of each industry, as well as the amount of each industry's output exported or used domestically for consumption or investment.

Our analysis uses the OECD Input-Output Database, which contains cross-sectional data on ten countries—the G-7 nations, plus Australia, Denmark, and the Netherlands—for selected years between 1968 and 1990.[17] These ten countries account for about two-thirds of world GDP and more than one-half of world trade. The tables divide the world economy into thirty-five sectors, including twenty-four goods-producing sectors, of which twenty-two are manufacturing. The concentration on manufacturing sectors is important because they increasingly dominate world trade.[18] For each country, we focus on the goods industries: agriculture and mining, plus the twenty-two manufacturing industries.

The OECD data offer two major advantages. First, they include an "imported transactions" table for each country, which reports the fraction of one sector's inputs imported from another sector. Hence, because the tables provide data on imported inputs, gross production (as well as value added), and exports, we can calculate the amount of vertical trade for each industry, as well as for the country overall. Second, the data provide a consistent set of tables to facilitate comparisons across countries and over time.

The input-output tables do, however, pose an aggregation problem. Each industry produces many goods, but we measure the use of imported inputs and exports at the industry level, rather than for each good. Consider an extreme example of a potential problem: An industry produces just two goods. One good uses imported intermediate inputs but is not exported. The other good uses no imported inputs but is exported. In this case, there would be no vertical-specialization-based trade, yet at the industry level we would calculate a positive amount of such trade. However, suppose that the first good relied heavily on imported intermediate inputs and was heavily exported, and the second good used no imported inputs and was not exported. Then, at the industry level, we would underestimate the amount of vertical-specialization-based trade. Unfortunately, we do not know whether the former or the latter case is more common, nor do we know the quantitative significance of the "bias."

Large countries generally find it easier than small ones, for scale economy reasons, to retain production of every stage of a good. This rationale explains why the United States, Germany, and Japan have three of the four lowest vertical trade shares of total trade.

When we calculate vertical trade as a share of total trade for nine countries in our sample for all available years between 1968 and 1990, two patterns emerge (Chart 7).[19] First, for every country but Japan, vertical trade as a share of total trade has increased from the first to the last year. Using the most recently available years for each country, we calculate vertical trade in our ten-country sample to be about 14.5 percent of overall trade, up from about 12.0 percent in the earliest years for each country—a 20 percent increase in less than two decades.[20] Second, there is wide cross-country variation in the amount of vertical trade. Japan, the United States, and Australia, for example, have the least amount of such trade, only about 7 percent of total trade in the final year of our study period. By contrast, 34.7 percent of the Netherlands' total trade was vertical trade in 1986 (the last year for which we have Netherlands data).

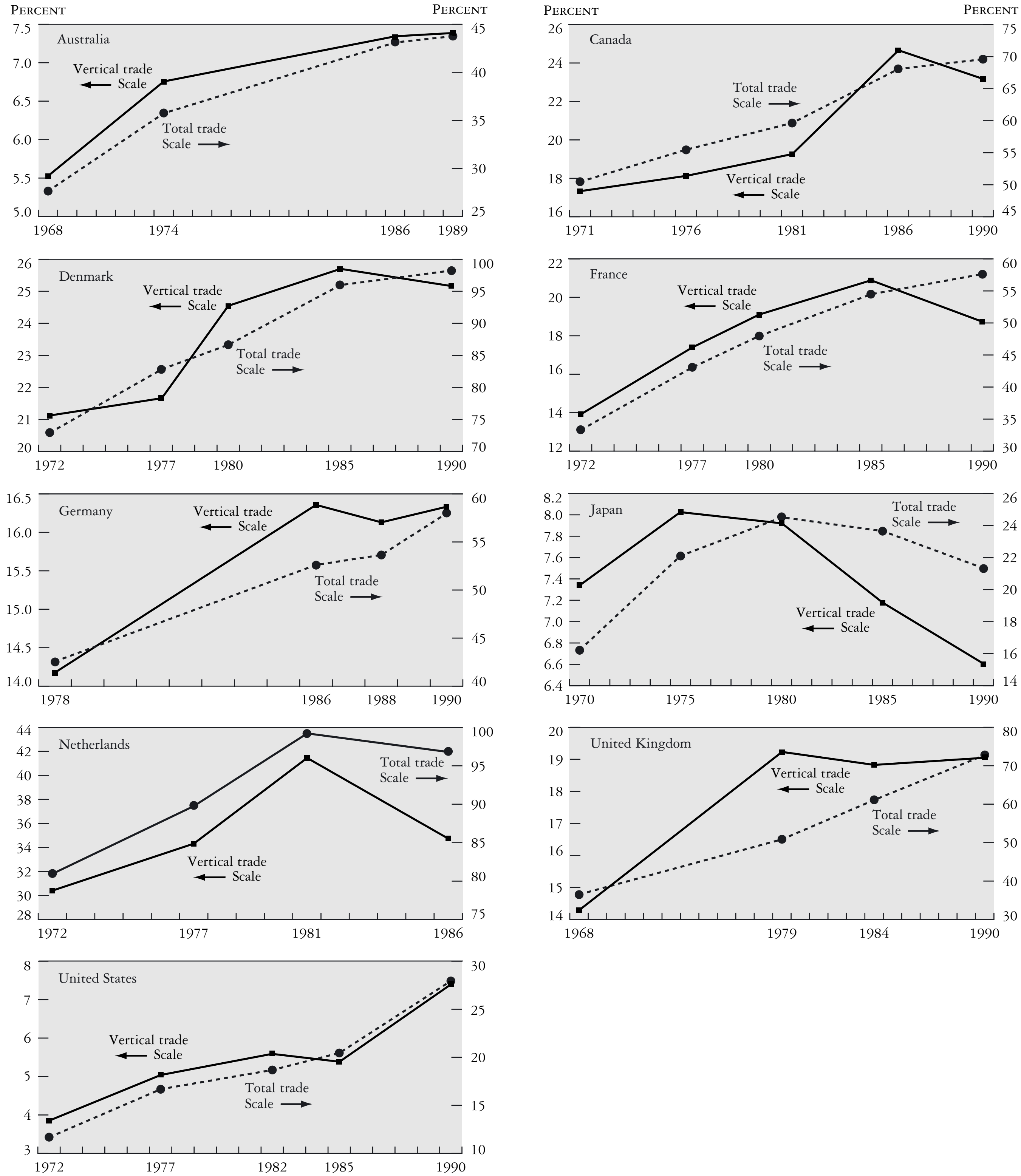
Interestingly, our ten-country sample includes the world's largest economies, which are conceivably the least likely to be involved in vertical trade. Large countries generally find it easier than small ones, for scale economy reasons, to retain production of every stage of a good. This rationale explains why the United States, Germany, and Japan have three of the four lowest vertical trade shares of total trade. The rest of the world, which accounts for about half of world trade, may be more like the Netherlands—which has much higher vertical trade. For example, in calculations not reported here, we find that vertical trade accounted for 28.4 percent of overall trade in Ireland in 1990 and about 23.9 percent in Korea in 1993. Moreover, our case studies suggest that countries like Spain, Malaysia, and the Philippines may also have large vertical trade shares. Consequently, a world vertical trade share on the order of 20 to 25 percent could well be likely.

Within each country, the prominence of vertical trade varies widely across industries (Table 1). Industries with the most vertical trade are motor vehicles, shipbuilding, and aircraft, as well as industrial chemicals, nonferrous metals, and petroleum and coal products; those with the least are agriculture, mining, wood products, and paper products. In Japan, for example, vertical trade accounts for 16 percent of industrial chemicals trade, while it accounts for only 0.1 percent of agriculture trade. There is also wide variation across countries within each sector. For instance, only 4.5 percent of motor vehicles trade in Australia is vertical trade, compared with 49.9 percent in Canada. Canada's figure is similar to the estimate calculated in our case study.

Our analysis also suggests that the increase in vertical trade is linked to the growing trade share of output. The positive relationship over time between vertical trade as a share of total trade and total trade as a share of gross merchandise output is evident in Chart 7. The correlation coefficient between the two variables for every country exceeds 0.79, except for Japan, where the correlation is 0.26.[21] To capture more formally the relationship between the trade share of output and vertical trade, we also perform

Chart 7

Vertical Trade as a Percentage of Total Trade and Total Trade as a Percentage of Gross Merchandise Output in Selected Countries



Source: Authors' calculations, based on the OECD's Input-Output Database.

Note: Italy is not shown because data were available only for 1985.

an ordinary least squares regression of total trade as a share of gross merchandise output on vertical trade as a share of total trade for all countries. We include in our regressions country-specific dummy variables to capture differences due to such factors as country size, GDP per capita, and distance from other countries. The regression results are reported below with the standard errors of the coefficient estimates in parentheses (we do not report the coefficients on the country-specific dummy variables):

$$(4)\quad \textit{Total trade (share of gross merchandise output)} = 0.03 + 2.92 \times \textit{vertical trade (share of total trade)}$$

$$(0.03)\qquad(0.41)$$

$$\textit{adjusted } R^2 = 0.95.$$

We find that the coefficient on vertical trade is statistically significant at the 1 percent level. The coefficient estimate implies that an increase of 1 percentage point in vertical trade as a share of total trade is associated with an increase of 2.92 percentage points in total trade as a share of gross merchandise output, an economically significant amount.[22]

We also employ growth decompositions to assess which industries account for the increase in a country's vertical trade as a share of total trade. The overall growth in vertical trade as a share of total trade depends on two forces: the change in each industry's vertical trade as a share of industry trade and the change in each industry's share of total trade. (Appendix B gives an algebraic derivation of the growth decomposition formula.) We examine the contribution of different industries to the growth in vertical trade as a share of total trade for Canada, France, Japan, the United Kingdom, and the United States. We concentrate

Table 1

VERTICAL TRADE AS A PERCENTAGE OF TOTAL INDUSTRY TRADE IN SELECTED COUNTRIES

Industry	Australia 1989	Canada 1990	Denmark 1990	France 1990	Germany 1990	Italy 1985	Japan 1990	Netherlands 1986	United Kingdom 1990	United States 1990
CHEMICALS										
Industrial chemicals	5.4	14.4	27.6	24.5	22.7	20.8	15.7	49.3	24.8	9.3
Drugs and medicines	3.8	3.4	31.0	25.7	0.0	16.4	1.9	32.2	15.2	3.5
MACHINERY										
Nonelectrical machinery	2.6	12.2	27.4	16.7	15.6	25.2	6.6	25.1	20.3	6.7
Office and computing machinery	0.0	29.2	0.0	19.8	11.5	28.2	8.7	24.1	29.3	16.7
Electrical apparatus, not elsewhere categorized	3.6	9.5	23.1	18.6	13.6	24.5	12.0	41.2	19.3	7.2
Radio, TV, and communication equipment	3.3	29.2	29.9	12.6	0.0	17.8	9.1	0.0	20.8	7.6
Shipbuilding and repairing	8.3	15.4	41.8	29.6	26.4	29.0	7.3	38.6	18.6	8.4
Other transport	5.2	26.1	22.0	10.5	0.0	15.4	7.2	15.4	16.6	8.1
Motor vehicles	4.5	49.9	0.0	21.1	22.4	16.1	5.6	25.4	20.0	8.7
Aircraft	2.4	28.6	0.0	38.4	16.7	24.2	7.3	52.5	34.5	11.6
Professional goods	4.2	11.8	27.6	11.3	11.6	13.1	7.4	24.9	16.8	5.8
OTHER										
Agriculture, forestry, and fishing	9.4	8.8	20.4	10.9	3.7	4.3	0.1	13.0	6.1	4.8
Mining and quarrying	12.6	5.8	3.5	3.5	1.3	0.0	0.0	4.8	10.5	0.8
Food, beverages, and tobacco	7.4	7.9	25.9	12.4	15.1	10.2	0.9	36.8	10.1	4.6
Textiles, apparel, and leather	9.2	7.6	29.4	18.1	15.1	29.6	5.2	34.6	19.1	2.8
Wood products and furniture	5.8	12.9	32.8	9.5	11.4	18.6	1.3	14.3	5.9	3.9
Paper, paper products, and printing	3.0	12.2	16.2	12.2	18.6	13.9	4.2	21.0	12.2	7.2
Petroleum and coal products	11.4	34.7	15.3	25.6	13.4	46.1	11.1	98.7	12.5	15.5
Rubber and plastic products	2.6	14.4	32.6	29.3	19.2	32.3	8.6	32.1	19.9	5.7
Nonmetallic mineral products	1.4	5.9	20.0	8.1	9.9	20.8	6.6	13.9	11.3	3.0
Iron and steel	9.3	15.5	16.9	21.5	11.8	31.9	13.4	27.1	20.9	4.7
Nonferrous metals	14.2	28.7	19.5	36.0	27.0	13.2	10.3	0.0	24.7	12.2
Metal products	9.3	11.9	26.3	13.5	16.1	22.7	7.3	27.7	17.2	8.9
Other manufacturing	5.4	10.4	28.6	10.8	16.6	42.1	4.2	21.3	11.0	2.8
TOTAL GOODS	7.4	23.2	25.2	18.7	16.3	19.6	6.6	34.7	19.1	7.4

Source: Authors' calculations, based on the OECD's Input-Output Database.

on the chemical and machinery sectors because they have accounted for the majority of manufacturing export share increases from the 1970s to the 1980s (see Ishii and Yi [1997, Table A.6]).

The top half of Table 2 presents our growth decomposition results. (The bottom half reports the change in each industry's vertical trade share of total industry trade, for reference.) In every country except Japan, we find that machinery accounted for at least 65 percent of the increase in the overall share of trade that is vertical trade between the first and last year of our data sample. Chemicals accounted for a smaller fraction of this increase across the five countries. Overall, in every country except Japan, these two industries together accounted for more than 75 percent of the growth in vertical trade as a share of total trade. Even in Japan, these industries' vertical trade as a share of total industry trade increased.[23] These results support our contention that, by and large, the industries that account for overall export growth are the same ones that account for vertical trade growth.

Table 2

Industry Contribution to Growth in Vertical Specialization from First to Last Year of Sample: Selected Countries

Percent

Country	Chemicals	Machinery	Other	Total
Canada	6.7	72.0	21.4	100[a]
France	23.7	65.0	11.3	100
Germany	4.8	80.1	15.1	100
Japan	-40.9	-263.9	404.9	100[a]
United Kingdom	12.7	124.5	-37.1	100[a]
United States	8.7	68.6	22.7	100
Change in Vertical Trade as a Share of Total Industry Trade[b]				
Canada	127.9	34.9	25.0	33.7
France	65.3	63.9	7.4	34.5
Germany	3.9	30.6	5.5	15.2
Japan	50.2	106.1	-61.5	-10.1
United Kingdom	29.1	151.2	-21.5	33.5
United States	117.2	134.2	45.4	92.2

Source: Authors' calculations, based on the OECD's Input-Output Database.

Notes: A detailed explanation of these growth decompositions is found in Appendix B. The industries that make up the categories "chemicals," "machinery," and "other" are listed in Table 1.

[a]Row does not sum to 100 because of rounding.

[b]Totals are weighted averages of the figures in columns 1-3, where the weights are each industry's share of total trade.

We use growth decompositions to answer another question: How has the growth of vertical trade contributed to the growth of total exports? Our decompositions allow us to calculate the contribution of vertical trade relative to horizontal-specialization-based trade. This calculation shows the percentage of growth attributable to each type of specialization. (Appendix B provides more details on the decomposition.) We find that for Canada and the Netherlands, *almost 50 percent* of the growth of exports from the first to the last year in the sample is due to growth in vertical trade (Table 3). In Denmark, France, and the United Kingdom, growth in vertical trade accounts for more than 25 percent of export growth. Only in the United States, Australia, and Japan does growth in vertical trade account for a small fraction of export growth. Table 3 also presents the change in vertical trade as a share of gross merchandise output for each country. We see that because vertical trade is still a relatively small fraction of total trade, growth in vertical trade accounts for less overall export growth than does growth in horizontal trade. However, vertical trade's increasing importance explains why its contribution to total export growth exceeds its share of total trade in all countries except Japan.

By and large, the industries that account for overall export growth are the same ones that account for vertical trade growth.

Causes and Policy Implications of Vertical Specialization

We have shown that vertical-specialization-based trade is rapidly increasing as a share of total trade. While our analysis does not permit us to conclude that the growth in vertical trade is *causing* the growth in world trade, three of our findings indicate a tight link between the two patterns. First, vertical trade as a share of total trade and trade as a

Table 3
CONTRIBUTIONS OF VERTICAL TRADE AND HORIZONTAL TRADE TO CHANGE IN EXPORT SHARE OF GROSS OUTPUT FROM FIRST TO LAST YEAR OF SAMPLE: SELECTED COUNTRIES

	Vertical Trade as a Percentage of Gross Output (Exports Only)[a]					Percentage of Change Due to Increase in	
Country	First Year		Last Year		Change in Export Share of Gross Output	Vertical Trade	Horizontal Trade
Australia	1968	0.8	1989	1.6	0.06	13.4	86.6
Canada	1971	4.4	1990	8.1	0.08	43.7	56.3
Denmark	1972	7.7	1990	12.4	0.17	27.3	72.7
France	1972	2.3	1990	5.4	0.11	28.4	71.6
Germany	1978	3.0	1990	4.7	0.09	19.4	80.6
Japan	1970	0.6	1990	0.7	0.03	3.2	96.8
Netherlands	1972	12.3	1986	16.8	0.10	47.4	52.6
United Kingdom	1968	2.6	1990	6.9	0.15	29.6	70.4
United States	1972	0.2	1990	1.0	0.07	11.9	88.1

Source: Authors' calculations, based on the OECD's Input-Output Database.

Note: A detailed explanation of these growth decompositions is found in Appendix B.

[a]Because we are accounting for the export share of gross output, we divide vertical trade by 2 to obtain the amount of vertical-specialization-based exports.

share of gross merchandise output are highly correlated over time in our sample of OECD countries. Second, the industries that account for the increase in total exports as a share of GDP—chemicals and machinery—also account for the increase in vertical trade as a share of total trade. Third, for most of our sample countries, growth in vertical trade accounts for 25 percent or more of the growth in overall trade. These findings link the increased internationalization of production to the rising trade shares of GDP and, consequently, enhance our understanding of the globalization of goods and services flows.

These findings raise two key questions: What have been the causes of vertical-specialization-based trade? And what are its consequences, especially the trade policy implications? Although these questions merit a more rigorous examination than we afford them here, we can shed some light on them.

CAUSES OF VERTICAL SPECIALIZATION

Most economists agree that decreases in tariff and nontariff trade barriers, as well as improvements in communications and the transportation of goods, have led to increased world trade. These decreased trade barriers and improvements in "distance-reducing" technologies have enabled countries to specialize in goods that they can produce relatively more efficiently.[24] China, for example, has become relatively more adept at manufacturing consumer products, while the United States has focused more on manufacturing high-tech products such as airplanes and business computers.

This concept of specialization, however, is the traditional horizontal one, which emphasizes production and trade of goods made entirely in one country. Vertical specialization carries the notion of specialization further, describing a process in which countries acquire expertise in particular stages of production. For example, computer production requires a skill-intensive stage—designing and manufacturing the chips—and a labor-intensive stage—assembling the computer. Vertical specialization allows countries to unbundle these stages so they can focus on those activities in which they are relatively more efficient. The reductions in trade barriers and improvements in transportation and communications technologies have facilitated this multicountry production sequence and thus have led to increased vertical specialization.

This phenomenon leaves open the question, Why has vertical specialization grown more rapidly than horizontal specialization? One answer is that improvements in communications technologies may favor vertical trade. Advances in such media as faxes, phones, pagers, e-mail, and videoconferencing have made it easier for countries to coordinate and monitor production in diverse locations. The dramatic increases in trans-Atlantic cable capacity and the corresponding reductions in the cost of trans-Atlantic communications have also encouraged frequent interaction between firms in different countries.[25] Since the sequential production nature of vertical specialization requires intensive oversight and coordination of production, these technological advances would tend to benefit vertical-specialization-based trade more than horizontal-specialization-based trade.[26]

Our case studies point to another explanation for vertical specialization's rising importance. When a good crosses only one border, tariffs and transportation costs are incurred only once. When a good crosses multiple borders, as in vertical-specialization-based trade, even low tariff rates of 2 to 4 percent are magnified as they are repeatedly applied to the good-in-process. This multiple taxation results effectively in much higher rates of protection. Therefore, reductions in these tariff rates will spur vertical specialization more than they will horizontal specialization. For example, the tariff rates on automotive trade between the United States and Canada before the 1965 Auto Agreement were roughly 10 to 15 percent, which was not high by historical standards. Yet the elimination of these tariffs resulted in a sixfold increase in auto trade in just four years, and raised the share of total trade accounted for by vertical specialization from zero to 20 percent in six years. Our maquiladora case study also suggests that tariff reductions have had a proportionately greater effect on vertical trade than on horizontal trade. Tariffs were reduced on both sides of the border; in particular, U.S. tariffs were changed so that they were levied only on Mexico's value added. This action removed the multiple-border-crossing penalty and led to more vertical-specialization-based trade.

Policy Implications

One clear policy implication of vertical specialization is that further tariff liberalization could yield substantial gains. This is true even though tariff rates in the developed countries are already quite low, approximately 5 percent or less. Moreover, the gains would become greater as vertical specialization increased.[27]

A second implication is that the linkage between trade policy and foreign direct investment policy is likely to tighten. Recall that all of our case studies involve multinational firms engaging in vertically integrated foreign direct investment. Absent firms' ability to invest freely in foreign countries, vertical-specialization-based trade may not occur. Similarly, it might not make sense to open a country to increased FDI inflows without also liberalizing import and export barriers. The notion that trade liberalization and investment liberalization are complementary has recently been supported by the theoretical literature as well (see, for example, Markusen [1997]).

One other trade policy issue is worthy of further consideration. Recently, there has been controversy surrounding the value of regional trade agreements such as the North American Free Trade Agreement, as opposed to broader agreements under the auspices of the World Trade Organization (see, for example, Bergsten [1997] and Bhagwati [1997]). The main argument against regional trade agreements is that they often create barriers between participating and nonparticipating countries; hence, trade "diversion" could occur, in which participating countries specialize in producing goods that nonparticipating countries would produce in a completely free world market. In other words, the gains to participating countries may be exceeded by the losses to nonparticipating countries.

One clear policy implication of vertical specialization is that further tariff liberalization could yield substantial gains.

It would be interesting to know how the increased importance of vertical specialization would affect gain-and-loss calculations. For example, would regional trade agreements lead to more or less trade diversion? What would happen to the attractiveness of regional trade agreements relative to world trade agreements? We note that all of our case studies involve either regional trading agreements or arrangements. In each case study, a relatively low-wage country engages in final assembly and a relatively high-wage country engages in parts and components production. These regional agreements and arrangements clearly boost trade and produce gains for the participants. Whether these gains come at the expense of other countries is yet to be determined.

Conclusion

The rising international trade shares of GDP are probably the most commonly cited evidence of the globalizing world economy. In this article, we identify a deeper dimension of the rising trade shares: the increased importance of imported inputs in the production of goods that are exported—that is, vertical specialization. Our evidence from case studies and input-output tables points to large and increasing shares of trade that can be attributed to vertical-specialization-based trade. In some of the smaller countries examined, the shares of total trade represented by vertical trade approach 50 percent.

Hence, globalization has gone beyond just "more trade." The nature of trade has changed to the point where countries increasingly specialize in producing particular stages of a good, rather than making a complete good from start to finish. This vertical trade is also what links heightened international trade to greater international production. In all likelihood, the forces that have led to increased vertical trade—lower trade barriers and improvements in transportation and communications technologies—will continue. Thus, we can expect the importance of vertical trade to grow as the world economy heads into the twenty-first century.

Appendix A: Calculations and Data Sources for Case Studies of Vertical Specialization

United States–Canada Auto Trade

Before the auto agreement, most automotive trade consisted of engines and parts. We conservatively attribute this trade entirely to the repair market. Because this trade is not vertical-specialization-based, in the first step of our calculation we estimate trade in the repair market from 1965 to the present and subtract that amount from the raw trade figures. We calculate the ratio of U.S. parts imported from Canada (and the ratio of Canadian parts imported from the United States) to total U.S. auto and truck output in 1964, and then assume that the ratios stay constant over time. Parts trade in the repair market in future years can then be estimated by multiplying these ratios by U.S. auto and truck output in those years. We subtract these estimates from the actual parts trade figures; the difference is our estimate of the parts trade destined for auto assembly that can be attributed to the auto agreement.

Our trade data are obtained from the United Nations Statistical Division's Comtrade Database, except for parts trade data between 1982 and 1994. Here, we use BEA figures because they include parts that are shipped for use in autos, such as air conditioners, but are counted by the United Nations in another, non-auto parts, category. The discrepancy between the United Nations and the BEA figures becomes significant only in the 1980s. Our "fraction of production exported" figures are obtained from *Ward's Automotive Yearbook* (1969-96).

United States-Mexico Maquiladora Trade

Our trade data come from the International Monetary Fund's *Direction of Trade Statistics* CD-ROM and from Banco de Mexico; the maquiladora data come from Instituto Nacional de Estadistica, Geografia e Informatica (INEGI).

Japan-Asia Electronics Trade

Our data are obtained from the Electronic Industries Association of Japan and the Japan Electronics Bureau.

Opel España Trade

Our value data on Opel España are obtained from the secretary-general of Opel España. Data on the number of cars produced and exported by all companies in Spain are obtained from the American Automobile Manufacturer's Association. Total Spanish auto trade data are obtained from the United Nations Statistical Division's *International Trade Statistics Yearbook* (1994).

Appendix B: Growth Decompositions

1. Industry contribution to growth in vertical trade as a share of total trade (Table 2):

Assume for simplicity that there are just two industries, C and M.

$VSBT_i$ = vertical-specialization-based trade in industry i. i = C,M

$HSBT_i$ = horizontal-specialization-based trade in industry i. i = C,M

$TT_i = VSBT_i + HSBT_i$ = total trade in industry i. i = C,M

$TT = TT_C + TT_M$ = total trade in the country

$VSBT = VSBT_C + VSBT_M$ = total VSBT in the country

ΔZ = growth in Z from initial year to final year

$$\frac{VSBT}{TT} = \frac{VSBT_C}{TT_C} \times \frac{TT_C}{TT} + \frac{VSBT_M}{TT_M} \times \frac{TT_M}{TT}$$

(industry contribution to VSBT as share of total trade)

$$\Delta\frac{VSBT}{TT} = \Delta\left(\frac{VSBT_C}{TT_C} \times \frac{TT_C}{TT}\right) + \Delta\left(\frac{VSBT_M}{TT_M} \times \frac{TT_M}{TT}\right)$$

(industry contribution to growth in VSBT as share of total trade)

(industry C's contribution is the first term on the right-hand-side of the equation divided by the left-hand-side of the equation)

2. Contribution of vertical trade and horizontal trade to change in export share of gross output (Table 3):

VSBT = vertical-specialization-based trade

HSBT = horizontal-specialization-based trade

X = total merchandise exports

Y = overall gross merchandise output

ΔZ = growth in Z from initial year to final year

Overall export share of gross output = X/Y

Vertical specialization export share of gross output = (VSBT/2)/Y

We divide VSBT by 2 because we are looking only at exports, not total trade.

Horizontal specialization export share of gross output = HSBT/Y = (X – VSBT/2)/Y

$X/Y = (VSBT/2)/Y + HSBT/Y$

$\Delta(X/Y) = \Delta((VSBT/2)/Y) + \Delta(HSBT/Y)$

Contribution of vertical specialization = $\Delta((VSBT/2)/Y) / \Delta(X/Y)$

ENDNOTES

The authors thank Lucinda Vargas-Ambacher and Tomoko Mischke for providing data and information on the maquiladoras and the Japanese electronics industry, respectively, and Jim Harrigan and two anonymous referees for very helpful comments.

1. Many economists, including Krugman (1995) and Irwin (1996), have noted that the international trade shares today are not much higher than they were in the early twentieth century. Nevertheless, the growth in trade is striking, and these economists acknowledge that the *nature* of trade is different today.

2. Balassa (1967, p. 97) may have coined the term vertical specialization. We later show how our adapted definition differs from his.

3. These case studies draw from Ishii and Yi (1997).

4. Balassa's definition of vertical specialization encompasses parts 1 and 2 of our definition. We also choose to include a part 3 to distinguish vertical specialization from intermediate goods trade in a broad sense. Sanyal and Jones (1982), among others, note that most imported goods—even so-called final goods such as motor vehicles—need sales and marketing services "added on" to them. Thus, almost all imported goods can be viewed as intermediate goods. By examining those goods that involve more than one border crossing, we limit our study to goods destined for export that are actually created through the sequential contributions of different countries. We thereby avoid the more ambiguous notion of intermediate goods.

5. Take Nike as an example. By most definitions, Nike is not a vertically integrated multinational firm because the footwear production occurs through arm's-length relationships. Yet to the extent that the footwear-producing countries import Nike services and other inputs and export Nike footwear, vertical-specialization-based trade occurs. In addition, vertical integration and vertical FDI deal with issues of ownership and internalization; vertical specialization does not.

6. In 1989, chemicals and allied products, machinery, and transportation equipment accounted for about 60 percent of manufacturing multinational gross product and about 35 percent of total multinational gross product (see Mataloni and Goldberg [1994]).

While multinational firms account for a majority of U.S. trade, their share of U.S. trade declined from 1977 to the mid-1990s. Zeile (1993, 1995) shows that the importance of foreign multinational firms to U.S. trade has been increasing; nevertheless, overall U.S. and foreign multinational trade has still been declining. These facts indicate that vertical trade goes beyond multinational firms.

7. Hereafter, "vertical-specialization-based trade" and "vertical trade" are used interchangeably.

8. While each good likely requires numerous production steps, we assume that these steps can be grouped into two stages.

9. See Economic Council of Canada (1975, p. 197). The agreement included two important restrictions: total production in Canada roughly had to match total sales in Canada, and 60 percent of the value added in Canadian-made cars had to be of Canadian origin (Wonnacott and Wonnacott 1967). A plausible argument can be made that the absence of these restrictions would have led to more vertical-specialization-based trade.

10. U.S. vehicles, engines, and parts exported to Canada as a fraction of total exports to Canada increased from 13 percent in 1964 to 30 percent in 1968. U.S. vehicles, engines, and parts imported from Canada as a fraction of total imports from Canada increased from less than 3 percent in 1964 to about 30 percent in 1968. As of 1995, engines and parts accounted for about 40 percent of U.S.-Canadian automotive trade. Total U.S. trade in vehicles, engines, and parts relative to U.S. auto and truck output increased from 9 percent in 1960 to 61 percent in 1994. Engines and parts accounted for more than 45 percent of total automotive trade in 1994.

11. Indeed, one of the major undergraduate textbooks in international economics, Krugman and Obstfeld (1997), discusses the auto agreement as such.

12. U.S. firms own the vast majority of maquiladoras, although there is increasing ownership by firms from Japan, Korea, and some European nations.

13. These are Harmonized Tariff System (HTS) items 9802.00.60 and 9802.00.80. They were formerly known as items 806.30 and 807.00 of the Tariff Schedule of the United States (TSUS). Item 9802.00.60 concerns tariff treatment for metal of U.S. origin processed in a foreign location and returned to the United States, while item 9802.00.80 involves goods that contain U.S.-made components (Hufbauer and Schott 1992, p. 93).

14. Much of the data that follow originate from Instituto Nacional de Estadistica, Geografia e Informatica (INEGI). Hanson (1996) draws from these data as well.

15. Over the last decade, two provisions have been passed to ease restrictions and one to tighten restrictions on the amount of maquiladora output that could be sold in Mexico. While there are no hard figures on the results of these rule changes, reports of factory managers in Mexico suggest that virtually all production is still exported to the United States. See Wilson (1992, pp. 40-1).

Note 15 continued

The presence of non-U.S.-owned firms in the maquiladora industry has made it likely that some of the inputs imported by Mexico are from non-U.S. sources. In 1989, approximately 4 percent of maquiladoras were Japanese- or Korean-owned. Moreover, it is plausible that U.S.-owned firms also rely on inputs imported from non-U.S. sources. However, we do not know the extent of this non-U.S. sourcing. In vertical trade calculations for Mexico, the origin of the imported inputs does not matter; it matters only when we calculate the vertical trade share of total United States–Mexico trade.

16. Vertical trade = \$3.6 billion + (\$3.6 billion) × (.7/.9) × (.78/.22) = \$13.5 billion. Total auto trade data are not yet available for 1995.

17. The G-7 nations are Canada, France, Germany, Italy, Japan, the United Kingdom, and the United States.

18. In 1970, manufacturing accounted for about 60 percent of world merchandise trade; in 1996, it accounted for about 75 percent.

19. Italy is not shown on the chart because the OECD data had information only for 1985.

20. We convert all countries' figures into 1989 dollars using 1989 exchange rates and the U.S. consumer price index. In several countries, vertical trade as a share of total trade declined between the next-to-last year and the last year. Most of the declines were relatively small, and none affects the underlying upward trends. By examining the data more carefully, we find that the largest decline, which occurred in the Netherlands between 1981 and 1986, is mainly explained by developments in the petroleum and coal products industry. The decline in oil prices between 1981 and 1986 would have lowered the share of imported inputs in gross output, reducing the estimated importance of vertical trade. Correlations presented below indicate that despite the decline, vertical trade as a share of total trade and trade as a share of gross output are highly correlated for all countries except Japan.

21. Some caution should be used in interpreting each correlation coefficient because there are only four or five data points for each country.

22. When we ran the regression in first differences and without the country dummies, we obtained similar coefficient estimates; also, the adjusted R^2 was 0.35.

23. The figures for Japan in the top and bottom halves of the table can be reconciled as follows: machinery vertical trade as a share of machinery trade grew, which made the fall in overall vertical trade as a share of total trade smaller, thereby contributing negatively to the (negative) growth.

24. These two forces have also facilitated specialization by allowing countries to take advantage of economies of scale. However, there is some disagreement about the relative importance of these two forces in explaining the growth of trade. See, for example, Rose (1991), Krugman (1995), and Bergstrand (1996).

25. See Hummels (1997).

26. Another factor that would favor vertical trade is the changing technology of goods production. Goods may require more production steps today than in the past: for example, the typical pharmaceutical drug needs fifteen or more production steps. The increase in the number of these steps heightens countries' opportunities to specialize in particular stages.

27. Ishii and Yi (1997) develop a model in which the gains from tariff reduction are several times larger when vertical specialization is included.

REFERENCES

Balassa, Bela. 1967. TRADE LIBERALIZATION AMONG INDUSTRIAL COUNTRIES. New York: McGraw-Hill.

Beigie, Carl. 1970. THE CANADA-U.S. AUTOMOTIVE AGREEMENT. Quebec: Canadian American Committee.

Bergsten, C. Fred. 1997. "American Politics, Global Trade." ECONOMIST, September 27, 23-6.

Bergstrand, Jeff. 1996. "The Growth of World Trade: Tariffs vs. Technology." Unpublished paper, University of Notre Dame, October.

Bhagwati, Jagdish. 1997. "Fast Track to Nowhere." ECONOMIST, October 18, 21-3.

Economic Council of Canada. 1975. LOOKING OUTWARD: A NEW TRADE STRATEGY FOR CANADA.

Hanson, Gordon H. 1996. "Localization Economies, Vertical Organization, and Trade." AMERICAN ECONOMIC REVIEW 86, no. 5: 1266-78.

Hufbauer, Gary C., and Jeffrey J. Schott. 1992. NORTH AMERICAN FREE TRADE: ISSUES AND RECOMMENDATIONS. Washington, D.C.: Institute for International Economics.

Hummels, David. 1997. "Towards a Geography of Trade Costs." Unpublished paper, University of Chicago.

Irwin, Douglas A. 1996. "The United States in a New Global Economy? A Century's Perspective." AMERICAN ECONOMIC REVIEW 86, no. 2: 41-6.

Ishii, Jun, and Kei-Mu Yi. 1997. "The Growth of World Trade." Federal Reserve Bank of New York Research Paper no. 9718, May.

Krugman, Paul R. 1995. "Growing World Trade: Causes and Consequences." BROOKINGS PAPERS ON ECONOMIC ACTIVITY, no. 1: 327-77.

Krugman, Paul R., and Maurice Obstfeld. 1997. INTERNATIONAL ECONOMICS: THEORY AND POLICY. 4th ed. New York: Addison-Wesley.

Markusen, James R. 1997. "Trade Versus Investment Liberalization." Unpublished paper, University of Colorado, August.

Mataloni, Raymond, Jr., and Lee Goldberg. 1994. "Gross Product of U.S. Multinational Companies, 1977-91." SURVEY OF CURRENT BUSINESS, February: 42-63.

Rose, Andrew. 1991. "Why Has Trade Grown Faster than Income?" CANADIAN JOURNAL OF ECONOMICS 24, no. 2: 417-27.

Sanyal, Kalyan K., and Ronald W. Jones. 1982. "The Theory of Trade in Middle Products." AMERICAN ECONOMIC REVIEW 72, no. 1: 16-31.

United Nations Conference on Trade and Development. 1996. WORLD INVESTMENT REPORT, 1996: INVESTMENT, TRADE AND INTERNATIONAL POLICY ARRANGEMENTS. New York: United Nations.

United Nations Statistical Division. 1994. INTERNATIONAL TRADE STATISTICS YEARBOOK. Vol. 1, TRADE BY COUNTRY. New York: United Nations.

U.S. Department of Commerce. Bureau of Economic Analysis. 1994-96. SURVEY OF CURRENT BUSINESS. Various issues. Washington, D.C.: U.S. Government Printing Office.

Ward's Communications, Inc. 1969-96. WARD'S AUTOMOTIVE YEARBOOK. 31st-58th eds.

Wells, Louis T. 1993. "Mobile Exporters: New Foreign Investors in East Asia." In K.A. Froot, ed., FOREIGN DIRECT INVESTMENT. Chicago: University of Chicago Press.

Wilson, Patricia A. 1992. EXPORTS AND LOCAL DEVELOPMENT: MEXICO'S NEW MAQUILADORAS. Austin: University of Texas Press.

Wonnacott, Ronald, and Paul Wonnacott. 1967. FREE TRADE BETWEEN THE UNITED STATES AND CANADA: THE POTENTIAL ECONOMIC EFFECTS. Cambridge: Harvard University Press.

Zeile, William J. 1993. "Merchandise Trade of U.S. Affiliates of Foreign Companies." SURVEY OF CURRENT BUSINESS, October: 52-65.

———. 1995. "Foreign Direct Investment in the United States." SURVEY OF CURRENT BUSINESS, May: 57-81.

THE FEDERAL RESERVE BANK OF NEW YORK
ECONOMIC POLICY REVIEW

Single copy subscriptions to the ECONOMIC POLICY REVIEW (ISSN 0147-6580) are free. Multiple copies are available for an annual cost of $12 for each additional subscription. Checks should be made payable in U.S. dollars to the Federal Reserve Bank of New York and sent to the Public Information Department, 33 Liberty Street, New York, NY 10045-0001 (212-720-6134). Single and multiple copies for U.S. subscribers are sent via third- and fourth-class mail.

ECONOMIC POLICY REVIEW subscribers also receive the Bank's Annual Report.

Library of Congress Card Number: 77-646559