New England Economic Review

Federal Reserve Bank of Boston

March/April 1991

The **New England Economic Review** is produced in the Research Department and is edited by Joan T. Poskanzer. Heather Schofield coordinates the design and production. The views expressed are those of the individual authors, and do not necessarily reflect official positions of the Federal Reserve Bank of Boston or the Federal Reserve System. The authors will be glad to receive comments on their articles. The Review is available without charge. Requests to be placed on the mailing list or for additional copies should be sent to the Research Library–D, Federal Reserve Bank of Boston, P.O. Box 2076, Boston, MA 02106-2076.

Contents

Contents

Regional Economic Conditions and the FOMC Votes of District Presidents

Geoffrey M. B. Tootell

Economist, Federal Reserve Bank of Boston. *The author wishes to thank Richard Kopcke and Stephen McNees for all their comments. Valuable research assistance was provided by Jeffrey B. Liebman.*

It is often argued that the institutional structure of the Federal Reserve System influences the formulation and attainment of national monetary policy goals. Havrilesky (1987, 1990), Havrilesky and Gildea (1990), and Belden (1989), for example, assert that Presidential appointment of members of the Board of Governors produces monetary policy sensitive to current political conditions. Conversely, the same literature, going back to Clifford (1965), also maintains that the power and independence of District Bank presidents remove them from accountability. As a result, some describe the Bank presidents as "too conservative," while others depict them as too sensitive to regional economic conditions. The first of these descriptions of Bank president behavior is examined in Tootell (1990b). The latter characterization is investigated here: does the decentralization of the Federal Reserve System (Fed) produce national monetary policy overly responsive to regional performance? This article quantifies the influence of regional conditions on District Bank voting by analyzing the monetary policy actually advocated by individual members of the Federal Open Market Committee (FOMC). The results indicate that District Bank presidents set policy dependent on national, not their regional, conditions.

District Bank presidents do play a major role in the formulation of monetary policy. The Federal Reserve Bank of New York always has one of the twelve votes at the policy-making FOMC meetings, and four of the remaining eleven votes rotate among the other Bank presidents. The enfranchisement of the Bank presidents is historical, perhaps an attempt to allay regional fears that monetary policy would accommodate the interests of the money center banks, although this article will suggest other reasons why the regional structure endures. Empowering regional institutions, however, could breed internecine squabbling at FOMC policy meetings. Whenever significant deviation between local and national interests occurs, or, in other words, when large variation exists in the economic performance of different regions, the potential for

regional conflict at the FOMC exists. Tootell (1990a) and Rosengren (1990) reveal the extent of the disparity in regional economic performance. The imperfect correlations between regional and national conditions permit an exploration of whether local interests play a disproportionately important role in the determination of Bank presidents' FOMC voting.

This article tests whether regional economic performance excessively influences the votes of District presidents. To explore the issue effectively, section I contains detailed analysis of why Bank presidents might vote on the basis of regional economic conditions. Because the empirical issues are complicated, a model of FOMC member behavior is carefully articulated. Section II discusses why the data used and the methodology chosen best examine the policy intentions of the FOMC. Section III presents the evidence concerning the effect of regional economic conditions on Bank president voting. A variety of tests consistently rejects the hypothesis that regional economic performance determines District policy-making. In fact, the results may highlight certain benefits to the current FOMC structure. A conclusion is provided in section IV.

I. Bank Presidents and the Regional Economy

Economic theory advises that regional economic conditions should have no effect on national policy independent of their impact on national performance. It is inefficient to use a national instrument, such as monetary policy, solely to influence regional economic performance. As an example, assume that real income in Region A declines while the rest of the country is fully employed. If the Fed eases policy, income may rise in A, but national inflation increases. Using a national instrument to affect a local target is like shooting a fly with a howitzer; one may be rid of the insect, but the collateral damage can be extensive. Employing a tool with less general effects, like a government works project in that area, would raise Region A's output without producing economywide inflation. Efficient use of policy, therefore, dictates that regional data should not affect Bank president FOMC votes outside of their influence on the national numbers. The national data correctly weight the importance of the regional variables in the economywide objectives of monetary policy, and thus, the votes of Bank presidents should only depend on the national figures.

A simple economic model can be applied to analyze the choices the FOMC makes. Since several subtle but important characteristics of this model will be highlighted, care is given to its exposition. All FOMC members are assumed to maximize their utility,

$$\underset{MP}{\text{Max}}\ U(\dot{Q}^E, \dot{P}^E), \tag{1}$$

by selecting the optimal monetary policy, MP, subject to two constraints,

$$\dot{Q}^E, \dot{P}^E = f_1(MP_t, \Omega_t) \tag{2}$$

and

$$MP = f_2(\text{Institutional Constraints}), \tag{3}$$

where $\dot{Q}^E$ is the expected growth in real activity, $\dot{P}^E$ is expected inflation, and Ω_t represents all the information possessed by the FOMC member at the time policy is determined. Utility is assumed to be a function of the policymaker's expectation of output growth and inflation; it depends on the expectations of these variables since monetary policy affects future, not current, conditions.[1] If other variables are of concern to the policymaker, they can also be included in the utility function. FOMC members choose the monetary policy that makes them as well-off as possible.

However, two constraints affect their policy selection. Equation 2 describes how the policymaker formulates his or her expectations of the goals. The expectations of output growth and inflation are the FOMC member's best forecasts of these variables given both the information they possess at the time and the monetary policy they initiate. The f_1 function translates MP_t and Ω_t into these best estimates. In essence, f_1 is the policymaker's model of the economy; it transforms the values of known economic variables, such as the money supply, interest rates, wages, fiscal policy, and the like, into future inflation and output growth. In fact, this model may differ among FOMC members; for a given monetary policy, two different policymakers with identical information sets can expect different values of future inflation and output growth if they possess different f_1s.

The second constraint represents the institutional factors that may limit policymakers' choices. For example, political pressures might affect FOMC votes. Either the Congress, which created the Federal Reserve System, or the Administration, which appoints Board governors, could influence the behavior of the FOMC.

Equation 3 introduces potential institutional costs to selecting the monetary policy that the FOMC finds optimal when equation 2 is the only constraint.

Although theory suggests that Bank presidents should not set policy contingent on regional economic conditions independent of their effect on national performance, the above paradigm illustrates three broad explanations for why they might. Bank presidents may care disproportionately about regional economic performance, thus injecting regional variables into their utility functions. Alternatively, the information sets, or the models, of the Bank presidents may be overly weighted toward regional variables. And finally, equation 3 emphasizes that Bank presidents may have institutional constraints on their behavior that could depend on regional conditions. Each explanation is discussed in detail below. Note, however, that the same methodology could be used to analyze Board governors. Regional variables would probably not be relevant in that case, but different models, f_1s, or institutional constraints might be important.

The procedure for appointment of District Bank presidents, as well as other aspects of the institutional structure of the regional Banks, could produce a constraint, represented by equation 3 in the above model, that binds District presidents to vote based on regional conditions. The Board of Directors of the District Bank plays an important role in selecting the Bank's president, even though the Board of Governors must approve the nominations. Furthermore, the Directors, in conjunction with the Board, also determine the president's salary. In fact, they also recommend changes in the discount rate. Although Bank presidents have considerable latitude in their FOMC voting, they wish to maintain good relations with their local Board. The Board of Directors consists of three local bankers, three local businessmen, and three other citizens. Their interests are generally more closely connected to regional economic conditions than to national performance.[2] It is, therefore, easy to conceive of the regional concerns of the Directors influencing the president's choice of monetary policy. If the District Boards do care inordinately about the local economy and Bank presidents' obligations to their District Boards do produce a different opportunity set, District presidents might vote dependent on regional economic performance.

On the other hand, Bank president utility functions could contain expectations of regional activity as well as the national goals included in equation 1. These utility functions may be skewed toward regional conditions because the directors who nominate Bank presidents might be biased toward individuals with such concerns. A prerequisite for nomination might be that one care disproportionately about local economic performance. Perhaps a more intuitive explanation for excessive regional concerns, however, relies on the local nature of the experiences and relationships of the Bank presidents. Their continual contact with the local community could make them overly sympathetic to their region's predic-

If presidents' votes are excessively sensitive to regional performance, the resulting monetary policy is suboptimal.

ament.[3] If presidents' votes are excessively sensitive to regional performance, because of differences in their constraint sets or utility functions, the resulting monetary policy is suboptimal.

Finally, regional variables may influence District Bank votes because their information sets or their models include a large share of regional data. Information sets and models are examined together because the two are so difficult to disentangle. The importance of the Ω_t can best be illustrated by examining the behavior of Bank presidents in relation to Board governors. Assume each District president has an information set before the FOMC meeting that differs from the information sets of the other presidents and that of the Board governors. Bank presidents could conceivably possess superior information on their regional economy because of both their knowledge of variables that are not aggregated into national figures and the lags that occur in data collection and aggregation. Conversely, Board governors could be more closely in tune with the national numbers since the large Board staff meticulously compiles and analyzes these data. If the two groups share information perfectly during the FOMC meeting and they possess identical models, f_1s, they will have the same expectations of output growth and inflation. Identical information sets, along with the same models and tastes, would produce identical votes. However, if they imperfectly share information, their votes could diverge even though their goals and constraints are identical. Monetary policy is inefficient, but

the root of the problem is the imperfectly shared information.

Not only might all FOMC members not possess the same information, they also need not possess the same model. In fact, certain District Banks are associated with certain paradigms. These various models, or f_1 functions, may emphasize different pieces of information and produce different expectations of the same goals. One would think, however, in the long run, different models could not be a source of divergent votes. If, for example, one regional Bank's model is a superior predictor of future GNP growth and inflation, then the FOMC should eventually recognize and share it. That model would then be vital to all FOMC members, and their forecasts would tend to converge.[4] Furthermore, any reliance of a Bank model upon a regional indicator beyond its ability to predict national performance would be inefficient. If regional economic developments do not help forecast the variables of interest to the Fed, or the FOMC members do not believe that they do, then these variables would have to be included in the utility functions of the Bank presidents in order to be significant in determining FOMC votes. Utilizing the model outlined in equations 1 to 3, the following two sections scrutinize the data to examine whether regional variables do influence Bank presidents' voting.

II. The Data and Methodology

Examining FOMC votes is both the only way to separate the monetary policy intentions of Board governors from those of Bank presidents, necessary when testing the effect of regional variables, and the superior method to analyze the determinants of that policy. The traditional reaction function literature, in McNees (1986), Havrilesky (1987), and Alesina and Sachs (1988), for example, measures the response of an assumed Fed intermediate target, such as the federal funds rate or the money supply, to assorted variables, like the growth rate of real GNP and the rate of inflation. Yet, using the movement of an assumed Fed intermediate target to proxy for the intentions of monetary policy not only fails to distinguish between the policy advocated by regional Bank presidents and Board governors but also creates serious problems in interpreting policy intentions.

In the context of the model outlined in section I, this methodology would substitute equation 3′ for 3, as monetary policy intentions are not directly measured but only approximated by the behavior of an assumed intermediate target, I_t.

$$I_t = f_3(\text{structural equations, institutional constraints, shocks, . . .}). \quad (3')$$

In the original model, changes in monetary policy were due to shifts in tastes or constraints. As discussed in Abrams, Froyen, and Waud (1980), unexpected movements in the Fed's intermediate target can, however, occur for reasons not associated with the variables of concern to the Fed. Because the "structural equations" for this instrument can change, one cannot be certain that monetary policy has shifted when the intermediate target that proxies for monetary policy intentions moves. The value of the intermediate target can change without a change in monetary policy intentions, and monetary policy can shift without movement in I_t.

An example best illustrates this side effect of using an intermediate target as a yardstick of monetary policy. Assume I_t is the money supply. An anticipated and accommodated movement in money demand unrelated to the variables of concern to the Fed will alter the money supply without changing Fed intentions. This shock affects I_t through the f_3 function in equation 3′, yet monetary policy remains constant. The traditional reaction function analysis interprets the change in money supply as either expansionary or contractionary monetary policy. In fact, during the 1970s and early 1980s institutional changes in banking caused unexpected movements in money realizations.[5] Since the relationship between the level of the instrument and the course of monetary policy is, in reality, unstable, the methodology represented by (3′) makes impossible any investigation of monetary policy intentions, as well as

Examining FOMC votes is both the only way to separate the monetary policy intentions of Board governors from those of Bank presidents and the superior method to analyze the determinants of member votes.

any attempt to disentangle the various roles constraints and tastes play in the formation of monetary policy.

The measure of the intention of monetary policy used in this paper does not suffer from the problems encountered by the intermediate target proxy. The actual vote of each member of the FOMC, whether for policy loosening, tightening, or no change, is recorded in the policy directives issued six weeks after every FOMC meeting. In these directives, FOMC members clearly discuss their votes in the context of their goals, and, in fact, the directives are fairly clear about when policy and intermediate targets diverge. FOMC votes, therefore, allow the separation of the movement in the intermediate target from the monetary policy intentions of each member of the FOMC. For this reason, the goals in equation 1, $\dot{Q}^E$ and $\dot{P}^E$, do not include any intermediate targets. Money, for example, is absent from the utility function in equation 1 even though many are "concerned" about it, because it is only an instrument to attain the goals of high GNP growth and price level stability. Counting the votes that dictate the direction of policy, not the movement of some intermediate target of the Fed's, winnows out most of the structural fluctuations unrelated to changes in policy intentions.[6]

Furthermore, directly measuring the votes of the FOMC members avoids the problem of deciding which intermediate target the Fed is using. If the incorrect target is examined, then no inferences about the direction of monetary policy are possible, no matter how stable the tool. This issue is discussed in detail in Luckett and Potts (1978, 1980) and Tootell (1990b). Since FOMC votes indicate the direction of policy, not which tool is used to accomplish that directive, this methodology circumvents another difficulty found in the traditional literature.

Accordingly, this article investigates the effect of current and expected economic conditions on FOMC voting by estimating the influence of certain explanatory variables, many examined in the traditional reaction function literature, on the probability of voting for tightening, loosening, or no change in policy.[7] The ability to distinguish between the votes of Bank presidents and Board governors also allows an examination of whether regional variables affect the probability of District presidents voting for certain policies. The regional indicators used here are economic data for each Fed district. If these variables significantly affect the votes of the District Bank presidents, then regional variables are influencing their monetary policy, whether because of different constraints or different utility functions.

Deciphering the intentions of policy from the directives is, of course, somewhat subjective. Although disagreements could arise over any one member's intentions at a particular meeting, the directives are usually clear; thus, votes are as good a proxy for members' policy desires as exist. Over the 1965–85 sample, approximately 58 percent of the votes cast were for no change in policy, 25 percent were for tightening, and 17 percent were for loosening. The large proportion of votes for no change illustrates the deliberate nature of monetary policy; the Fed was not constantly attempting to fine-tune the economy. Furthermore, episodes of tightening and loosening tended to be clumped together, indicating a "gradualism" in policy tightening or easing. As Brainard (1967) points out, doubts about the magnitude of the effect of changing policy provide a justification for relatively small policy moves at one time; the Fed preferred a series of slight policy changes in the same direction to a large single jump.

The "Green Book," circulated to FOMC members before each meeting, contains the Board staff's expectations of the future paths of output growth, inflation, and unemployment.

Both actual and forecasted data are used to examine the effect on FOMC voting of national and regional variables. Unfortunately, testing the effect of regional conditions on Bank president voting is inhibited by the dearth of economic data collected at the state level. This study aggregated the statewide data that were available into Fed District figures.[8] Contemporaneous values of regional and national employment growth rates and unemployment rates, national inflation rates, and dummies for the deviations in growth of real per-capita gross regional product were used in the analysis.[9] Alternatively, since the lags in the effects of Fed policy require the FOMC to react to its expectations about future economic variables, the

sensitivity of FOMC voting to forecasts of national unemployment, real GNP growth, and inflation were also examined. The "Green Book," which is circulated to FOMC members before each meeting, contains the Board staff's expectations of the future paths of output growth, inflation, and unemployment. The contemporaneous data provide proxies for the regional economic conditions while the Green Book forecasts furnish good measures of the Board staff's expectations of the national variables.[10]

Because policy decisions depend so heavily on the exact information available at the time, care must be given in each estimation procedure to the timing of the data. The frequency of the explanatory variables found in the Green Book exactly coincides with the FOMC votes, and these forecasts are updated before each FOMC meeting. The independent variables that are not found in the Green Book, however, have a frequency different from the Green Book forecasts and the FOMC votes. Employment growth, for example, is released monthly, while the FOMC has met from eight to twelve times a year over the sample period investigated in this article. To account for this problem, the most recent employment growth figure available before the FOMC meeting is used as an explanatory variable for that meeting's vote. The tests in this paper are constructed to ensure that the independent variables contain the most recent additions to each member's information set.

Table 1
Regional Effects on the Simple Model

	(1) All FOMC Members	(2) All FOMC Members	(3) Board Only	(4) Banks Only	(5) Banks Only
Tightening					
C	−2.241	−2.371	−2.75	−2.023	−2.145
	(13.21)	(13.65)	(11.21)	(8.07)	(8.38)
$\dot{Q}^E$	.136	.078	.038	.087	.027
	(8.07)	(4.11)	(1.33)	(3.32)	(.88)
$\dot{P}^E$	.163	.157	.172	.135	.129
	(7.30)	(7.01)	(5.58)	(4.12)	(3.92)
$\dot{RL}$		.118		.08	.028
		(7.01)		(4.19)	(1.21)
$\dot{USL}$			.213		.164
			(6.63)		(4.06)
Loosening					
C	.175	.190	.176	.206	.215
	(.97)	(1.04)	(.73)	(.72)	(.74)
$\dot{Q}^E$	−.211	−.176	−.129	−.208	−.138
	(11.79)	(8.59)	(4.34)	(6.87)	(3.85)
$\dot{P}^E$	−.177	−.176	−.164	−.190	−.186
	(6.12)	(6.04)	(4.27)	(4.18)	(4.02)
$\dot{RL}$		−.057		−.030	.028
		(3.12)		(1.31)	(1.00)
$\dot{USL}$			−.115		−.167
			(3.65)		(3.50)

Note: t statistics in parentheses. $\dot{Q}^E$ is the forecast of real output growth and $\dot{P}^E$ is the forecast of inflation. $\dot{USL}$ is the rate of national employment growth, $\dot{RL}$ the regional rate. $\dot{RL}$ in equation (2) uses the District employment growth rate for Bank presidents and the national rate for the Board governors.

III. The Results

This section begins with the empirical results of the simple model described in section I. The effects of regional variables on District president voting are then examined. A variety of indicators of District conditions are tested and all reject any effect of regional variables on Bank voting. Next is shown the robustness of the result when adding other national variables, consistent with the past work on Fed reaction functions. Complicating the voting function in no way alters the rejection of regional effects. Finally, the coefficients are interpreted as functions of tastes and constraints.

Regional Variables in the General Model

Table 1 presents the results using multinomial logit estimation for the simplest model in section I. The equations in Table 1 are derived from the utility function given in the model; the FOMC sets monetary policy in order to attain its output growth and inflation targets. The first equation in Table 1 provides the coefficients for the effect on the probability of voting to tighten (top panel) and loosen (bottom panel) relative to a vote of no change, given the Green Book forecasts of the change in real GNP and inflation.[11] The coefficients on all these forecasts are of the expected sign and are statistically significant. As expected real GNP growth increases, the probability of voting for tightening rises while the probability of voting for loosening declines.[12] Similarly, an increase in expected inflation raises the probability of tightening and decreases the probability of loosening. This result supports the hypothesis that the Fed attempts to maintain a balance between output growth and inflation.

As employment growth is one of the best indi-

cators of regional economic conditions, it is the first component of regional information tested. Equation (2) of Table 1 adds a regional employment growth rate to the first equation in Table 1. Since this equation examines the FOMC as a whole, the employment growth rate uses the regional employment for each District Bank and national employment for the Board of Governors. The coefficients on all the variables in the second equation are statistically significant and of the anticipated sign. When employment growth rises, the probability of voting for tighter policy rises and for looser policy falls. Similarly, when the expected inflation rate increases, the probability of tightening rises and the probability of loosening declines. The magnitudes of the coefficients on the Green Book forecasts of real GNP growth decrease, but they remain statistically significant. Again, these coefficients are consistent with the belief that the Fed attempts to balance off the performance of the real economy, proxied by employment growth, and inflation.

The employment data suggest that Bank presidents vote on national, not regional, variables.

The significance of the employment variable is of central concern to this article. Equation (2) in Table 1 suggests the potential importance of regional variables. However, since the regional employment variable for the FOMC as a whole combines the District employment growth faced by the Bank presidents with the national employment growth faced by the Board governors, the finding of a significant effect on voting for this variable could be due to the overwhelming significance of the national employment figures for the Board governors. To test this hypothesis, the regression is divided into a Board equation (3) and a Bank equation (4). Although the magnitude of the regional employment growth coefficients fall slightly when examining the Bank equation, (4), versus the FOMC equation, (2), the coefficient for loosening is correctly signed and statistically significant while that for tightening is correctly signed.[13] An increase in employment growth in a given Fed District increases the probability of that District's president voting for tighter policy and decreases the probability that he or she will vote for loosening. The significance of the regional employment growth coefficient for the Bank presidents seems to suggest that District Bank presidents do vote based on regional economic conditions.

However, once the national employment growth rate is accounted for in the Bank presidents' voting function, the regional variable loses all significance. Including the national employment growth in the Bank presidents' equation, equation (5) of Table 1, tests whether the regional variable helps explain District Bank voting beyond its correlation with its national counterpart. Both the statistical significance and the magnitude of the coefficient on regional employment growth collapse, while the coefficient for national employment growth is statistically significant and large. Although the regional and national employment growth rates are far from perfectly correlated, they move together sufficiently so that the change in regional employment was given credit for the change in national employment when the national rate was omitted in equation (4); the regional employment growth in equation (4) merely captures the effect of changes in national employment on Bank president voting. In fact, the employment data suggest that Bank presidents vote based on national, not regional, variables.

Robustness Using Other Measures of Regional Activity

Other proxies for regional economic performance are also investigated to ensure that regional conditions are being captured effectively. It is possible that District employment growth is not the best indicator of local conditions. Therefore, the effect of District unemployment rates, although available in sufficient length only for the larger states, is tested under the identical procedure used in Table 1.[14] The pattern of results using regional and national unemployment rates is similar to that using regional and national employment growth; although the regional unemployment rate shows some statistical significance in an estimation including the Green Book forecasts of real GNP growth and inflation alone, when the national unemployment rate is included, the regional figure loses all significance. Again, regional unemployment is only acting as a proxy for the movement in the national number.

Another possible measure of regional conditions is a dummy variable derived from an estimate of Gross District Product. The effect on voting of deviations of the Fed District product from its trend

growth rate is examined.[15] Again, this different measure of District conditions basically replicates the results using regional employment growth. Equations consisting of this dummy variable and the Green Book forecasts of real GNP growth and inflation find slight significance for the dummy, but all the significance disappears when this variable's national counterpart is included. Thus, the rejection of the importance of regional economic conditions is extremely robust to different proxies for regional economic conditions.

Robustness of Results Using Other National Variables

The omission of other variables of interest to the Fed could affect the outcomes of these tests; inclusion of these other variables might reverse the finding of insignificant regional effects. These variables, mentioned throughout the traditional reaction function literature, could be other goals of the Fed, or other important indicators besides the Green Book forecasts. For example, some District Banks are believed to emphasize the growth of the monetary aggregates more than others. These Banks may believe that the money supply is a better predictor of long-run inflation, and thus weight it more heavily than the other FOMC members. Although an exhaustive test of all the variables hypothesized to be of interest to the Fed is beyond the scope of this article, the robustness of the rejection of regional concerns is examined for different specifications for the Fed's behavior.

The Fed may be concerned with real variables besides output growth. Table 2 contains the coefficient estimates of the basic voting function, equation (1) of Table 1, when first the national employment growth rate, then the Green Book forecast of the change in the unemployment rate, and finally both variables are added to the regression. Using either of these two measures clearly reduces the size of the coefficient on the Green Book forecast of GNP growth. And, when all three are included in the same equation, both the forecast of GNP growth and the forecast of unemployment rate changes tend to lose significance and importance. Essentially the three variables are attempting to measure the same thing, the Fed's concern about the performance of the real side of the economy; as a result, each variable's effect on voting is difficult to disentangle from the others'. For this reason only one such measure of real activity, the Green Book forecast of real GNP growth, will be used subsequently.[16]

Tests were performed to ensure that the rejection of the regional variables is robust to the use of these other proxies for the FOMC's national goals. In equation (5) of Table 1 the importance of regional employment is rejected when both national employment growth and real GNP forecasts are considered. In fact, when the GNP forecasts are dropped from that equation, regional employment still produces no statistically significant effect on Bank president voting; when only inflation expectations and regional employment are considered, regional employment is important, but when national employment is added to the equation the regional variable loses all significance. Furthermore, replacing the Green Book forecast of output growth in equation (1) of Table 1 with its prediction of national unemployment rate changes produces the identical result; the change in regional unemployment is somewhat significant when only it and inflation forecasts are included in the regression,

Table 2
Alternative Targets of Real Activity

	(1) All FOMC Members	(2) All FOMC Members	(3) All FOMC Members
Tightening			
C	−2.464	−2.154	−2.594
	(13.95)	(11.67)	(13.09)
$\dot{Q}^E$	.032	.115	.057
	(1.53)	(4.56)	(2.14)
$\dot{USL}$	.202		.212
	(8.69)		(8.74)
Δune^E		−.265	.336
		(1.25)	(1.46)
$\dot{P}^E$	.152	.160	.154
	(6.76)	(7.19)	(6.83)
Loosening			
C	.202	−.428	−.239
	(1.05)	(1.97)	(1.07)
$\dot{Q}^E$	−.133	−.07	−.052
	(5.83)	(2.21)	(1.62)
$\dot{USL}$	−.125		−.092
	(5.10)		(3.55)
Δune^E		1.33	.962
		(5.18)	(3.55)
$\dot{P}^E$	−.173	−.161	−.164
	(5.87)	(5.43)	(5.51)

Note: *t statistics in parentheses.* $\dot{Q}^E$, $\dot{P}^E$, *and* Δune^E *are the Green* Book forecasts of output growth, inflation and change in the unemployment rate. $\dot{USL}$ is the growth rate of national employment.

but regional unemployment loses all significance when its national counterpart is included. The previous results are, therefore, robust to alternative specifications of the national goals of monetary policy. Using national employment growth or unemployment expectations as Fed goals rather than real GNP growth still produces a rejection of the hypothesis that District Bank presidents are overly concerned about regional performance.

Other variables frequently hypothesized as important to the Fed were also examined. For example, some of the traditional reaction function literature includes money supply growth since it is believed to be an intermediate Fed target. The Fed attempts to achieve some rate of money growth in order to attain its goals for GNP and inflation. Basically, the Fed may be thought to use the rate of growth in the money supply as an accurate indicator of future nominal output growth. Yet, for money growth to be important in equations that include the Green Book forecasts of inflation and output growth, FOMC members must either believe money growth is not adequately considered in these forecasts, or care about horizons different from the Green Book forecasts and believe that money growth helps predict GNP growth and inflation at these different horizons, or care about money for its own sake. Furthermore, the traditional literature has also hypothesized that the Fed is concerned with smoothing interest rates. The lagged change in the federal funds rate might be an important determinant of FOMC votes and is, therefore, also added to the basic voting equation.[17] The FOMC equations are used to test the importance of money growth and lagged changes in the interest rate since they are national, not regional, variables.

Equation (1) of Table 3 includes the rate of growth in the money supply. It is statistically significant and of the expected sign; when money growth is high (low) the probability of tightening (loosening) increases. The coefficient is, however, quite small, between one-seventh and one-tenth the magnitude of the other coefficients. Equation (2) in Table 3 adds the lagged change in the federal funds rate to equation (1) of that table.[18] The coefficients for this variable are large and also statistically significant. The probability of tightening (loosening) rises (falls) after a recent change in the federal funds rate.

Yet, rather than capturing concerns over interest rate smoothing, the lagged federal funds rate is probably illustrating certain time series properties of FOMC policy. The Fed tends to tighten or loosen slowly; that is why similar votes are clumped together. The lagged change in the federal funds rate is merely a surrogate for last month's monetary policy. In fact, when a dummy variable of various lags was used to capture lagged policy, the change in the federal funds rate lost all significance and magnitude.[19] Both the money growth rate and the lagged change in the federal funds rate are statistically significant in Table 3, but the added information they provide appears less important; the coefficient on money is small relative to the other coefficients, and the lagged federal funds rate simply seems to be catching the serial correlation of policy votes.

What is essential to this study, however, is that neither variable affects the coefficients for the original equation. The Green Book forecasts of inflation and output growth remain essentially the same when these additional variables are included. Running the identical test of the regional effects using equation (2)

Table 3
Alternative National Targets

	(1) All FOMC Members	(2) All FOMC Members
Tightening		
C	−2.439	−2.535
	(12.88)	(12.66)
$\dot{Q}^E$	.139	.132
	(8.18)	(7.29)
$\dot{P}^E$	.170	.110
	(7.53)	(4.4)
$\dot{M}$	.022	.06
	(2.44)	(6.05)
ΔFF		.921
		(8.82)
Loosening		
C	.335	.329
	(1.74)	(1.68)
$\dot{Q}^E$	−.21	−.150
	(11.68)	(7.90)
$\dot{P}^E$	−.179	−.179
	(6.20)	(6.01)
$\dot{M}$	−.024	−.063
	(2.50)	(5.43)
ΔFF		−1.21
		(10.51)

Note: t statistics in parentheses. $\dot{Q}^E$ and $\dot{P}^E$ are the Green Book forecasts of output growth and inflation. $\dot{M}$ is the growth rate of M1. ΔFF is the lagged change in the federal funds rate.

of Table 3 as the base equation again rejects the hypothesis that regional performance affects FOMC voting. Even when that test is performed for Bank presidents only, the hypothesis is rejected. Using any of the possible specifications in this article, it appears that District presidents did not vote dependent on regional employment growth.

As a final test, contemporaneous data at the quarterly frequency are examined. Reducing the frequency is an attempt to capture the most recent trends in economic conditions while filtering out the noise in the monthly data. Quarterly data also account for the inertia inherent in monetary policy. Various lag structures were studied using the contemporaneous data, since the timing of the information with quarterly observations is more complicated.[20] Table 4 includes employment growth and Green Book inflation forecasts; it is basically a quarterly version of Table 1. In fact, the results in Table 1 are completely replicated. The volatility of the employment figures is clearly not driving the rejection of the regional variables. The more slow-moving quarterly data produce the same results as the perhaps noisier FOMC frequency figures.

Table 4
Contemporaneous Quarterly Data

	(1) All FOMC Members	(2) Banks Only	(3) Board Only	(4) Banks Only
Tightening				
C	−1.19	−1.02	−1.41	−1.19
	(19.25)	(12.29)	(14.40)	(11.65)
$\dot{RL}$	.625	.33		−.155
	(3.88)	(1.70)		(.63)
$\dot{USL}$			1.169	1.181
			(4.12)	(3.11)
$\dot{P}$	.758	.714	.783	.675
	(4.87)	(3.13)	(3.66)	(2.93)
Loosening				
C	−1.11	−1.33	−.908	−1.14
	(19.54)	(14.99)	(12.00)	(12.30)
$\dot{RL}$	−2.289	−1.276		−.277
	(12.71)	(7.34)		(.90)
$\dot{USL}$			−3.06	−3.012
			(11.08)	(6.95)
$\dot{P}$	−.687	−.973	−.446	−.769
	(3.66)	(3.28)	(1.84)	(2.57)

Note: t statistics in parentheses. $\dot{USL}$ is the growth rate in national employment. $\dot{P}$ is the acceleration of actual inflation. $\dot{RL}$ in equation (1) uses the District employment growth rate for Bank presidents and the national rate for the Board governors.

Models versus Constraints

Can judgments be made concerning the tastes of the FOMC members from the results in this paper? Discerning tastes is essential if judgments are to be made about the course of policy or the origin of possible side effects of certain appointment procedures. In the previous literature, tastes could not be discerned. As the model in the first section of this article highlights, it is uncertain whether changes in policy are driven by tastes or constraints. For example, if expected GNP growth increases, members of the FOMC could vote to tighten policy either because they do not care much about output growth yet have a tremendous distaste for inflation or because they believe any excess output growth will greatly increase inflation. In the first case their tastes are revealed, in the second their models. Under special circumstances, however, these tastes can be distinguished. If the FOMC members' actual expectations are known, and the FOMC members are not constrained as in equation 3 of the model, the results using these expectations would reveal their tastes. Model differences would not affect the coefficients, as the actual expectations would already incorporate the models that produced them. Thus, if the Green Book forecasts actually are the FOMC members' expectations of real GNP growth and inflation, their coefficients provide information on the relative tastes of the FOMC for these two variables.[21]

On the other hand, if the Green Book forecasts do not perfectly represent FOMC members' expectations of these variables, then their tastes cannot be so easily discerned. To examine this issue, actual GNP growth and inflation minus their Green Book forecast values are added to the regression in Table 1. If these two new variables are statistically significant and of the expected sign, then the FOMC members seem to be adding to the expectations of the Green Book.[22] As seen in Table 5, the Green Book forecast error terms have the predicted effect on the votes to tighten and are almost always statistically significant. As actual GNP growth or inflation exceeded its Green Book forecast, the FOMC was more likely to tighten. The coefficient on the residuals for the real GNP growth for policy loosening is correctly signed, though insignificant. Only the coefficient on inflation for the probability of loosening is the wrong sign and statistically significant.

To ensure the robustness of the importance of

these errors, a slightly different methodology is also employed. An ordered probit is performed in which it is assumed that as real GNP growth or inflation decreases, the probability of a vote moves constantly away from tightening, through no change, and toward loosening; the coefficients are, thus, assumed to be constant.[23] The ordered probit results suggest that actual real GNP growth beyond the Green Book forecast has the expected effect on FOMC policy; members are more likely to vote for tightening (loosening) when GNP grows faster (slower) than the Green Book forecast. The inflation forecast error has no statistically significant effect in the ordered probit.

Table 5
Forecasts and Expectations

	(1) All FOMC Members	(2) All FOMC Members
Tightening		
C	−2.26	−2.54
	(13.15)	(12.46)
$\dot{Q}^E$	.129	.124
	(7.47)	(6.67)
$\dot{P}^E$	.158	.107
	(7.00)	(4.19)
$(\dot{Q} - \dot{Q}^E)$	.063	.064
	(3.44)	(3.35)
$(\dot{P} - \dot{P}^E)$	.104	.111
	(2.94)	(3.01)
$\dot{M}$		.057
		(5.67)
ΔFF		.929
		(8.89)
Loosening		
C	.011	.197
	(.06)	(.98)
$\dot{Q}^E$	−.221	−.164
	(11.81)	(8.21)
$\dot{P}^E$	−.160	−.172
	(5.47)	(5.67)
$(\dot{Q} - \dot{Q}^E)$	−.019	−.003
	(.82)	(.12)
$(\dot{P} - \dot{P}^E)$	.165	.147
	(4.36)	(3.79)
$\dot{M}$		−.056
		(4.77)
ΔFF		−1.17
		(10.17)

Note: t statistics in parentheses. $\dot{Q}^E$ and $\dot{P}^E$ are forecasts of output growth and inflation. $\dot{M}$ is the M1 growth rate and ΔFF is the change in the federal funds rate. $(\dot{Q} - \dot{Q}^E)$ and $(\dot{P} - \dot{P}^E)$ are the errors in the Green Book forecasts of output growth and inflation.

The importance of the real output forecast error seems robust; thus, the FOMC members were apparently bringing information beyond that contained in the Green Book.

One possible explanation for the significance of the prediction error is that different Banks bring different, perhaps superior, "models" to the meeting. In fact, different economic models are often associated with different District Banks. If diverse models are driving regional Bank behavior, then individual Banks should be reacting differently to the independent variables in this study. In Tootell (1990b) a variety of tests were performed on various equations comparing single Banks, and even a group of reputedly conservative Banks, with the remaining District Banks. The hypothesis that all Banks vote identically can almost never be rejected.[24] The similarity of the Banks is a somewhat surprising result considering the diversity of the paradigms associated with different Banks. Either the models are, in essence, not all that different, the votes of the Bank presidents are not dictated by these diverse paradigms, or a consensus and reconciliation is worked out in the FOMC meeting. The similarity among Banks also justifies an empirical assumption made throughout the article: since all Banks react alike, it is not a problem to lump them all together and constrain them to have the same coefficients in the above tests.

The hypothesis that all Banks vote identically can almost never be rejected.

Because of the uncertainty about the exact regional variable to test, many different ones are examined. Using all of these various regional indicators, the evidence strongly rejects the hypothesis that Bank presidents relied disproportionately on regional economic conditions. Yet the FOMC members did seem to bring important information to the meetings. Tootell (1990b) reveals the similarity in voting whether measured between different Banks or between Bank and Board; thus, this information was shared among all FOMC members. Information was brought to the table but it was not regional in nature and no regional disputes occurred over it.

IV. Conclusion

The conclusion that Reserve Bank presidents have different concerns than Board governors has become a common assumption. In one view, these presidents are trapped by their constituencies, the District Boards of Directors, who somehow constrain or influence the presidents to protect local interests. Yet, the evidence presented here indicates that presidents did not manipulate monetary policy to help their own regional economies. Either the directors took a broader view than the hypothesis asserts or the presidents were more independent than assumed. Comparing the FOMC votes to the discount rate votes would be one way to test which of these alternatives was more probable. A different view postulates that Fed Bank presidents are too politically independent, and thus, far more likely to be for tighter policy. Although some differences between Banks were perceptible, Tootell (1990b) used the methodology presented in this article to reject the hypothesis that Bank presidents were any more conservative, or prone to "tight" policy, than the Board governors. Since they voted the same, their choice of monetary policy was identical, both their models and their tastes were close to identical.

A consensus-forming tendency could be the force that drives out any differences in tastes or models. The improvement in the Green Book forecasts suggested in this paper was probably the result of information added by the interaction among FOMC members before the vote. No evidence has been found to support the contention that this information was regional. Furthermore, this added information was most likely shared as all members tended to vote alike. In fact, perhaps the ability to capture and utilize different information is the reason the regional diversity endures at the Fed. Yet, the exact appointment procedure, or institutional structure, does not seem to affect the voting behavior of FOMC members. Assumed differences within the Fed cannot be used as a reason to alter the institution.

---

[1] For ease of exposition, the policymaker is maximizing utility with respect to the expected values of inflation and output growth instead of the expected utility of the actual realization of these variables. These two approaches are equivalent if the monetary authority is risk neutral. Altering the approach does not change any of the analysis essential to this paper.

[2] The nine directors of each District Bank will, in general, have much closer ties to the regional economy. Businesses whose products are nationally distributed are certainly less apt to be disproportionately concerned with local conditions, but producers of non-traded goods and providers of regional services will be overly dependent on regional economic performance.

[3] A rigorous example of such a situation would model an altruistic president. His or her utility would depend on those in closest proximity, those he or she has the most contact with. As a result, the president would overly weight the utility of local residents versus the rest of the country and tend to vote dependent on regional performance.

[4] The change of operating procedures and instruments through the 1970s and 1980s signifies a change in models dominating the Fed.

[5] In "The Case of the Missing Money" Goldfeld (1976) examines in detail the extent of the unexpected shortfall in money demand produced by any traditional money demand function before that time.

[6] Dissents are sometimes made in FOMC voting for technical reasons. The explanations for these technical dissents are articulated in the minutes of the directives. Maintaining the example in the body of the paper, if only one FOMC member believes the money demand function has shifted for reasons unrelated to changes in income or inflation, he or she might make a technical dissent for changing the money supply while keeping policy constant. Thus, these dissents were not included as disagreements with policy in this paper.

[7] Qualitative analysis is the method used for estimation. As there are three alternative policy responses, the results shown in the paper derive from multinomial logit procedures. The coefficients represent the change in the probability of choosing either to tighten or loosen relative to the choice of no change in policy at given values of the independent variables. Ordered probits were also performed and are mentioned only on the rare occasion when they do not corroborate the logit results.

[8] All data were aggregated using a weighted average. For example, employment growth was weighted by the state's share in total District employment. One complication, however, is that several Fed Districts include parts of states. These states were incorporated into the District that contained the larger share of that state. This problem is not serious as the divided states are usually quite small relative to each District.

[9] The employment data were seasonally adjusted using the Census X-11 procedure. Inflation was measured using the Consumer Price Index (CPI). The Gross State Product numbers are published by the U.S. Bureau of Economic Analysis (BEA) and were aggregated, like the employment data, into District figures.

[10] The sample periods on the two data sets do not perfectly coincide. The contemporaneous sample ranges from 1963–86, while the Green Book data cover 1966–85. The Green Book sample is shorter because the forecasts were not begun until the mid-1960s and because the data are not available until five years after a given FOMC meeting. When both types of measures are used in the same estimation procedure, the sample is constrained by the shorter Green Book period.

[11] All results presented here use the Green Book's one-quarter-ahead forecast. The results are basically identical when the two-quarters-ahead forecast is used.

[12] Various measures of the output variable were examined in various forms of the voting function. Deviations from de-trended

output, output growth with drift, and growth that included dummies for the post-1974 sample all produced essentially the same coefficients and significance levels.

[13] In regressions that drop the output growth variable in order to avoid the potential collinearity problem faced by including output growth and employment growth in the same equation, both regional employment growth coefficients are statistically significant for the Bank equations.

[14] Because unemployment rates are not available for all states over much of the sample, the state where the District Bank is located is used as the regional unemployment rate. This proxy permits a longer sample period. Unemployment levels and changes were examined, in an attempt to capture movements from some full employment level.

[15] The total Fed District product is derived from the BEA's gross state product series. The regional output is de-trended and the residuals are used; these deviations from trend derive from Tootell (1990a). As gross state product is an annual series, the value of this residual is the same over the entire year.

[16] That the lag of an actual, known, variable, the employment growth rate, tends to perform better than the Green Book forecast of real GNP growth suggests that forecasts are less important than actual recent figures. However, when testing measures of inflation this result is reversed; generally only the Green Book forecast of price level changes is significant, not actual, past inflation. Because of both the difficulties of interpreting coefficients when there is multicollinearity and the results using the inflation forecasts, the Green Book forecast of real activity is used in the remainder of the paper.

[17] The Fed might be concerned about the volatility of the bond market; thus, the FOMC will not allow the interest rates to change radically.

[18] The money growth rate is for M1. The lagged change in the federal funds rate is the monthly average of the month before the FOMC meeting.

[19] The dummy variable took the value of zero if the previous vote were to tighten, one if it were for no change, and two if the vote was to loosen. Including a single lag of policy decreased the size and significance of the lagged change in the federal funds rate by about 40 percent. Including two lags eliminated the coefficient and its significance completely.

[20] The complication over the timing of the lag structure is due to the uncertainty about information flows. For example, using fourth-quarter growth in real GNP to explain the first-quarter FOMC vote is suspect since information on output is received over the course of the entire quarter. However, using contemporaneous real GNP growth assumes information not yet officially received. Regressions with different lags of the independent variables were investigated and found to be fairly consistent whether contemporaneous or lagged values are used. Note this is only a problem with the contemporaneous quarterly data, not the Green Book forecasts.

[21] This statement is true if the utility horizon is the same as the Green Book forecasts. Also, risk neutrality is still being assumed. These assumptions allow one to factor out all the models, constraints, and the like, to get straight to the concerns of the FOMC members.

[22] It is possible that FOMC concerns over other variables, correlated with national conditions, could produce statistically significant forecast error coefficients. It is doubtful that tastes are the cause of these results, however. A large enough percentage of the FOMC would have to care about the same unusual variable for the coefficient on the entire FOMC to be significant. Secondly, the coefficient of any such variable, the stock market or the dollar for example, need not result in the correct sign of the voting coefficient; what exactly is driving movements of that variable can be negatively or positively correlated with national conditions, and this correlation can change given different circumstances or shocks. Furthermore, if this variable is an instrument, like the money supply or the federal funds rate, it is "cared about" by definition only in its relation to its targets, like GNP and inflation. Traditionally when one assumes some FOMC member "cares" too much about a variable, that member believes the Green Book or other forecasting tools are not considering the informational content of that variable sufficiently. This is simply a difference in models, not a difference in utility functions. Finally, including many of these variables, like the money supply and the interest rate, failed to remove the significance of the forecast error so the error was not picking up any utility effects from these variables. Thus, although the significance of the error is possibly due to a lucky correlation, it seems unlikely. The problem is the variable selected to be the instrument for expectations, not that the model is misspecified.

[23] This test is the only time the ordered probit produces a slightly different result than the multinomial logit. Using this methodology, the problem with the wrong signed coefficient on the inflation error disappears.

$$P(\text{loosen}) = \underset{(20.9)}{1.70} - \underset{(17.12)}{.125(\dot{Q}^E)} - \underset{(10.04)}{.11(\dot{P}^E)}$$

$$- \underset{(3.14)}{.027(\dot{Q} - \dot{Q}^E)} + \underset{(1.48)}{.023(\dot{P} - \dot{P}^E)}$$

The coefficient on the error in real GNP is correctly signed, so the probability of loosening decreases with a rise in the GNP growth rate forecast error, and is statistically significant. The coefficient on the inflation error term is insignificant.

[24] Only the Federal Reserve Bank of St. Louis comes close to a rejection when using the model in Table 2, while both Boston and St. Louis reject using the second equation in Table 3. The St. Louis Bank votes tended to depend on money more than the votes of the other Banks. The Boston votes depended more on money, forecast errors, and the lagged change in the federal funds rate than the others.

References

Abrams, Richard K., Richard Froyen, and Roger Waud. 1980. "Monetary Function Reaction Functions, Consistent Expectations, and The Burns Era." *Journal of Money, Credit, and Banking*, vol. 12, no. 1, February, pp. 30–42.

Alesina, Alberto and Jeffrey Sachs. 1988. "Political Parties and the Business Cycle in the United States, 1948–1984." *Journal of Money, Credit, and Banking*, vol. 20, no. 1, February, pp. 63–81.

Beldon, Susan. 1989. "Policy Preferences of FOMC Members as Revealed by Dissenting Votes." *Journal of Money, Credit, and Banking*, vol. 21, no. 4, November, pp. 432–41.

Brainard, William. 1967. "Uncertainty and the Effectiveness of Policy." *The American Economic Review*, May, pp. 411–25.

Clifford, A. Jerome. 1965. *The Independence of the Federal Reserve System*. Philadelphia, PA: University of Pennsylvania Press.

Goldfeld, Stephen M. 1976. "The Case of the Missing Money." *Brookings Papers on Economic Activity*, (3), pp. 683–730.

Havrilesky, Thomas M. 1987. "A Partisanship Theory of Fiscal and Monetary Regimes." *Journal of Money, Credit, and Banking*, vol. 19, no. 3, August, pp. 308–25.

———. 1990. "The Influence of the Federal Advisory Council on Monetary Policy." *Journal of Money, Credit, and Banking*, vol. 22, no. 1, February, pp. 37–50.

Havrilesky, Thomas M. and John Gildea. 1990. "Packing the Board of Governors." *Challenge*, vol. 33, no. 2, pp. 52–55.

Luckett, Dudley G. and Glenn T. Potts. 1978. "Policy Objectives of the Federal Reserve System." *Quarterly Journal of Economics*, August, pp. 525–34.

———. 1980. "Monetary Policy and Partisan Politics." *Journal of Money, Credit, and Banking*, vol. 12, no. 3, August, pp. 540–46.

McNees, Stephen K. 1986. "Modeling the Fed: A Forward-Looking Monetary Policy Reaction Function." *New England Economic Review*, November/December, pp. 3–8.

Rosengren, Eric S. 1990. "How Diversified Is New England?" *New England Economic Review*, November/December, pp. 3–16.

Tootell, Geoffrey M.B. 1990a. "Central Bank Flexibility and the Drawbacks to Currency Unification." *New England Economic Review*, May/June, pp. 3–18.

———. 1990b. "Are District Banks More Conservative Than Board Governors?" Mimeo, Federal Reserve Bank of Boston.

Stock Market Efficiency: An Autopsy?

Peter Fortune

Professor of Economics at Tufts University and Visiting Scholar at the Federal Reserve Bank of Boston. *The author is grateful to Richard Kopcke and Joe Peek for constructive comments, and to David Zanger for his able research assistance.*

This is the second in a series of three papers assessing the performance of the U.S. stock market. The first paper (Fortune 1989) dispelled the myth of increasing stock market volatility: it found that the monthly total rate of return on the Standard & Poor's 500 Composite Index has not been more volatile in the 1980s than in previous periods. Indeed, the peak of stock market volatility was in the 1930s. Others (for example, Schwert 1989) have reached the same conclusion using data going as far back as 1859. These observations suggest that investors do not face greater uncertainty about the returns on stocks than they have in the past, at least over periods of a month or longer. They also suggest that firms need not be concerned that the cost of equity capital has risen for risk-related reasons.

The present paper addresses the question of the efficiency of the stock market—do stock prices correctly reflect available information about future fundamentals, such as dividends and interest rates? Stated in another way, is the volatility of stock prices due to variation in fundamentals, or do other sources of volatility play a significant role?

The third paper will complete the trilogy by investigating the nature and consequences of very short-term (daily or intra-day) volatility. While the market's volatility over a period of a month or longer has not been increasing, rare—but prominent—daily spikes in stock price variation, which remain largely unexplained, have become the subject of public policy debate. Hence, the next paper will address episodes such as the October 19, 1987 crash and the break of October 13, 1989.

The present paper is structured as follows. Section I discusses the meaning of the Efficient Market Hypothesis (EMH) and draws out some of its implications for stock price behavior. The second section reviews the major stock market anomalies that cast clouds over the hypothesis of market efficiency, while the third section assesses "modern" evidence against the efficiency hypothesis. Section IV proposes an explanation for market inefficiency that is consistent with much of the evidence mar-

shalled in sections II and III. The paper concludes with a brief summary.

The purpose of this paper is not to draw out the policy implications of market inefficiency—that is the task of the next paper. However, the prominence of inefficiencies suggests a role for public policies that might be counterproductive in an efficient market. In short, this paper suggests that recent proposals for changes in margin requirements, introduction of trading halts, and other reforms might be productive.

I. The Efficient Market Hypothesis (EMH)

Practitioners are interested in the stock market because it is their bread and butter. Academic economists are interested for a very different reason: for them, the stock market provides an excellent laboratory for the evaluation of microeconomic theory. Common stocks are highly standardized products traded in an active auction market with very easy exit and entry of both producers (firms issuing equity) and consumers (investors purchasing shares); as a result, the prices of common stocks should conform to the implications of the theory of competitive markets.

The Efficient Market Hypothesis is the focus of the laboratory experiments, for it is the logical result of the application of microeconomic theory to the determination of stock prices. As Marsh and Merton point out (1986, p. 484):

> To reject the Efficient Market Hypothesis for the whole stock market . . . implies broadly that production decisions based on stock prices will lead to inefficient capital allocations. More generally, if the application of rational expectations theory to the virtually "ideal" conditions provided by the stock market fails, then what confidence can economists have in its application to other areas of economics where there is not a large central market with continuously quoted prices, where entry to its use is not free, and where short sales are not feasible transactions?

There were several reasons for the popularity enjoyed by the EMH in the 1960s and 1970s. First, it was rooted in a very strong theoretical foundation. This foundation began with Samuelson's work on the behavior of speculative prices, in which he showed that the prices of speculative assets should follow a random walk (1965). It was further buttressed by Harry Markowitz's theory of portfolio selection (1959) and by William Sharpe's construction of the Capital Asset Pricing Model (CAPM), which described the implications of optimal portfolio construction for asset prices in security market equilibrium (1964); both Markowitz and Sharpe won the 1990 Nobel Prize in Economics for their contributions. The final contribution was Robert Lucas's Rational Expectations Hypothesis (1978), which examined the implications of optimal forecasting for individual behavior and macroeconomic performance.

By the late 1970s, those who disputed the EMH found themselves facing an avalanche of sharply pointed and well-argued opposing positions. The theorists had apparently won, in spite of the paucity of supportive evidence, and the prevailing view was that no systematic ways exist to make unusual returns on one's portfolio. Practitioners were reduced to the position of ridiculing the EMH but could not make effective arguments against it.[1]

A second reason for the popularity of the EMH was the hubris of financial market practitioners during the 1960s. The "go-go" years had rested upon the notion that opportunities for unusual profits were abundant, and that it required only a reasonable person and a bit of care to sort out the wheat from the chaff in financial markets. The EMH provided an antidote to this hubris, for it argued that opportunities to make unusual profits were both rare and ephemeral: by their very nature, they were the result of temporary market disequilibria that are quickly eliminated by the actions of informed traders. Thus, the EMH counseled healthy skepticism in investment decisions. This skepticism about "beat the market" strategies has led to the popularity of index funds, which allow investors to hold "the market" without worrying about individual stocks.

"If the application of rational expectations theory to the virtually 'ideal' conditions provided by the stock market fails, then what confidence can economists have in its application to other areas of economics?"

The EMH, Definition 1: Prices Are Optimal Forecasts

The fundamental insight of the EMH is that asset prices reflect optimal use of all available information. A more formal statement is that the price of an actively traded asset is an optimal forecast of the asset's "fundamental value." To understand this notion, suppose that market agents think of each possible sequence of future events as a "state-of-the-world," and that there are N possible states of the world, to each of which a number s (s = 1,2,3, . . . , N) is assigned. For example, state-of-the-world 1 might be "dividends grow at 2 percent per year indefinitely, and a constant discount rate of 5 percent should be used," while s = 2 might be "dividends grow at 3 percent for two years, during which the discount rate is 7 percent, but thereafter dividends grow at 1 percent and the discount rate is 4 percent." Suppose also that the set of all the available information at time t is denoted by Ω_t and that $\pi(s|\Omega_t)$ is the probability that state s will occur, conditional on the information set available at time t.

Then for *each* state-of-the-world we can calculate a fundamental value of the asset, which we denote as P*(s). Hence, P*(4) is the fundamental value if state number 4 occurs. Note that there is no single fundamental value, rather there are N possible fundamental values, one for each state. If P_t is the market price and the *expected* fundamental value is $E(P_t^*|\Omega_t) = \Sigma_S P^*(s)\pi(s|\Omega_t)$, the EMH is embodied in the statement that the current price of the asset is equal to the expected fundamental value, or, more concisely, that the price of an asset is the best estimate of its fundamental value; that is,

$$(1) \qquad P_t = E(P_t^*|\Omega_t).$$

Consider the following simple example. There are three states-of-the-world: in state 1, the fundamental value is $100, in state 2 it is $75, and in state 3 the fundamental value is $40. Investors do not know which state will materialize, but they have formed an assessment of the probability of occurrence of each state. Suppose that these probabilities are 0.25, 0.35 and 0.40, respectively. The market price under the EMH will be the expected fundamental value: P_t = 0.25($100) + 0.35($75) + 0.40($40) = $67.25.

The EMH carries a number of strong implications about the behavior of asset prices. First, recall that Ω_t contains all the relevant information available at time t; this includes historical information (for example, past values of the asset price, the history of dividends, capital structure, operating costs, and the like), as well as current publicly available information on the firm's policies and prospects. Because P_t already incorporates all of the relevant information, the *unanticipated* component of the market price should be uncorrelated with *any* information available at the time the price is observed. A simple test of this proposition is to do a regression of P_t on a measure of the optimal forecast $E(P_t^*|\Omega_t)$ and upon any information that might be in Ω_t (say, the history of the stock price). This crude form of technical analysis[2] should result in a coefficient of 1.0 on $E(P_t^*|\Omega_t)$ and coefficients of zero on past stock prices, leading to the conclusion that all information in past stock prices has been embedded in the fundamental value.[3]

The fundamental insight of the EMH is that asset prices reflect optimal use of all available information.

A second implication bears on the sequence over time of prices under the EMH. Suppose we are at time t and we wish to forecast the price at time t + 1. If we knew the information that would be available at time t + 1, our forecast would be $E(P_{t+1}^*|\Omega_{t+1})$. But we do not know, at time t, the information available at time t + 1; we only know Ω_t. If r is the required rate of return on an asset with the risk level and other characteristics of the asset under consideration, the best forecast of P_{t+1} when we only know the elements in Ω_t is $E(P_{t+1}^*|\Omega_t)$.[4] Now, any new information arriving between time t and time t + 1 is, by definition, random so its effect on price creates a random deviation from today's best forecast.

From the optimal forecast definition of the EMH in (1), we see that the EMH implies the following sequence of prices:

$$(2) \quad P_{t+1} = (1 + r)P_t + \epsilon_{t+1}, \qquad \text{where } E(\epsilon_{t+1}) = 0$$

Equation (2) says that the sequence of prices will be a random walk with drift; price will vary randomly around a rising trend. Because new information is random, having no predictable components, ϵ_{t+1} has a zero mean and is without serial correlation.

This is the basis of the "random walk" tests of

the EMH which estimate equations like (2) and search for serial correlations in the residuals. Appendix 1 uses time series analysis to determine whether daily changes in the closing value of the S&P500 index during the 1980s are consistent with a random walk. The answer appears to be "almost, but not quite." The results show a five-day trading cycle that can be used to predict stock price movements, but this cycle is a very small source of the total variation in the S&P500. Hence, it might not be strong enough to generate economic profits after transactions cost; this is, of course, consistent with the EMH.

The EMH, Definition 2: Risk-Adjusted Returns Are Equalized

The EMH can be restated in a different manner, which focuses on the rate of return on individual assets rather than their prices. The Capital Asset Pricing Model (CAPM), developed by Markowitz and Sharpe, states that in an efficient market the risk-adjusted *expected* returns on all securities are equal; any differences across assets in expected rates of return are due to "risk premia" arising from unavoidable (or "systematic") uncertainty.

The CAPM distinguishes between two types of risk: systematic risk, which affects all securities, each to a different degree, and unsystematic risk, which is unique to individual securities. Unsystematic risk can be avoided by an appropriate diversification of portfolios, based on the variances and covariances of security returns. Because unsystematic risk can be avoided without any sacrifice in the return expected from the investor's portfolio, it imposes no risk premium.

However, systematic risk, which affects all securities and is, therefore, unavoidable, will earn a risk premium. The CAPM defines a simple measure of the amount of systematic risk a security contains: its "beta coefficient." A security's beta coefficient measures the marginal contribution of that security to the market portfolio's risk: if $\beta = 0$, adding the security to the optimal portfolio does not affect portfolio risk; if $\beta < 0$, portfolio risk is reduced by adding the security to the portfolio: and if $\beta > 0$, the security adds to the portfolio's risk. The returns on a security with a beta of 1.0 move with the market—when the return on the market portfolio changes by 1 percent, the return on a $\beta = 1.0$ security moves in the same direction by 1 percent. Securities with betas greater than 1.0 have above-average risks, while securities with betas below 1.0 have below-average risks.

According to the CAPM, the realized return on a security is described by the following "characteristic line":

$$R_i = r_f + \beta_i(R_m - r_f) + \upsilon_i \qquad (3)$$

where R_i is the realized return on the specific security, r_f is the return on a risk-free asset (such as U.S. Treasury bills), R_m is the rate of return on the market portfolio (such as the S&P500), and υ_i is a zero-mean random variable whose variance measures the unsystematic risk of that security. The slope of this characteristic line, β_i, is the security's beta coefficient. The beta can be estimated by a bivariate regression of the excess return on an asset, $(R_i - r_f)$, on the excess return on the market, $(R_m - r_f)$.

Equation (3) describes the relationship between the *realized* return on an individual security and the *realized* return on the market. This provides the basis for answering the question, "What is the normal return on a security?" From the characteristic line we can see that because $E(\upsilon_i) = 0$, the *expected* return on an asset will be a linear function of the asset's risk level, as measured by its beta coefficient. For every security, the expected return will lie on the same straight line, called the Security Market Line (SML), which relates expected return to risk (beta). The SML is described by the following equation:

The Capital Asset Pricing Model states that in an efficient market the risk-adjusted expected returns on all securities are equal; any differences across assets in expected rates of return are due to "risk premia" arising from unavoidable uncertainty.

$$E(R_i) = r_f + (E_m - r_f)\beta_i \qquad (3')$$

where $E(R_i)$ is the *expected* return on the *ith* security, and E_m is the *expected* return on the market portfolio. This relationship represents the optimal forecast of a

security's rate of return (rather than its price); under the EMH any deviations of the actual return, R_i, from this relationship must be random.

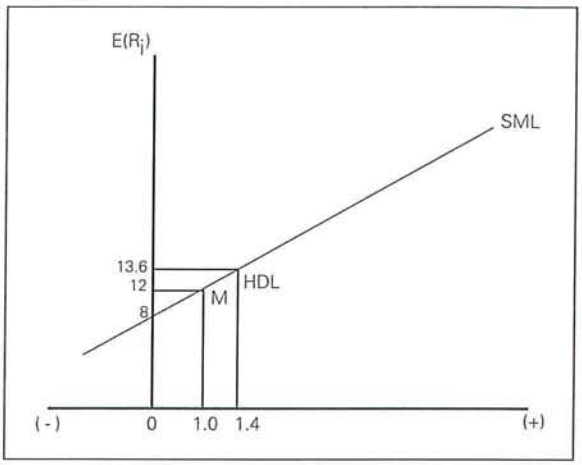
The term $(E_m - r_f)\beta_i$ is the "risk premium" for the *ith* security; it is the product of the market reward for a unit of risk, defined as the expected excess return on the market portfolio, and of the security's risk level, measured by its beta. The SML says that the optimal forecast of the return on a security is the risk-free interest rate plus a risk premium. Hence, on a risk-adjusted basis (when the risk premium is deducted from the expected return), all securities are expected to earn a return equal to the risk-free interest rate. Securities that have high betas, hence adding more to the portfolio risk, will carry higher expected returns because risk-averse investors will require higher returns on average to compensate them for the additional risks.

Each possible value of r_f and E_m will have a different SML describing security market equilibrium. Figure 1 shows the SML under the assumptions r_f = 8 percent and E_m = 12 percent. In this case the intercept will be 8 percent and the slope of the SML will be 4. The market portfolio would be at point "M," with a beta of 1.0 and an expected return of 12 percent; to verify this, simply substitute $\beta = 1.0$ into equation 3'. Discovery of a security whose *expected* return is above (or below) the SML is an indication of market inefficiency, for that security is expected to give a risk-adjusted return above (or below) the required level; that is, it is underpriced (or overpriced).

Thus, the SML can be used to describe the expected returns that are consistent with an efficient market. For example, point "HDL" on Figure 1 represents Handelman Corporation, a distributor of home entertainment media (records, video tapes, etc.). As of October 1990, Handelman's beta coefficient was 1.4, so the hypothetical SML predicts a return on HDL of 13.6 percent. Handelman's higher return is attributable solely to its above-average risk.

Figure 1

Efficient Markets: The Security Market Line



The equation for the SML is $E = r_f + (E_m - r_f)\ \beta_i$.
The SML drawn above assumes $r_f = 8\%$ and $E_m = 12\%$.

Some Caveats

Both forms of the EMH rest on a strong assumption: the market equilibrium of asset prices is independent of the distribution across investors of the two basic raw materials of investment: information and wealth. In short, all those things that make different investors evaluate assets differently are treated as of negligible importance. Among these "irrelevant" factors are differences in probability assessments (more optimistic investors will invest a larger share in favored assets), differences in transactions costs (investors with low costs, such as financial institutions, will devote more resources to stocks than will high-cost investors, such as individuals), and differences in tax rates paid by investors.

If these factors can be ignored, the prices or returns on assets will be determined solely by fundamentals. But if they are important, prices can deviate, perhaps persistently, from fundamental values. Indeed most explanations of inefficiency in security markets rest on some form of heterogeneity among investors.

II. Stock Market Anomalies and the EMH

The theoretical victories of the EMH were not supported by empirical evidence. True, some studies did support the EMH; for example, numerous studies showed that stock prices were random walks in the sense that past stock prices provided no useful information in predicting future stock prices. But these studies do not represent the preponderance of the evidence for two reasons. First, gross inefficiencies can coexist with random walks in stock prices, as in

the case of rational bubbles (which is discussed below). Second, and more important, by the 1980s a vast literature on stock market anomalies had developed. These anomalies, defined as departures from efficient markets that allow economic agents to enjoy unusually high (risk-adjusted) returns, appeared to lead to rejection of the EMH.

This section reviews some of the major anomalies in stock price determination that have been the traditional basis for rejecting the EMH. While a panel of coroners might not declare the EMH officially dead, by the late 1980s the burden of proof had shifted to the EMH adherents. The anomalies discussed in this section are among the list of causes that would appear on the death certificate.

The Small-Firm Effect

Arguably the best-known anomaly in stock prices is the Small-Firm Effect: the common stocks of small-capitalization companies have, on average, exhibited unusually high rates of return throughout most of this century. This is shown clearly in Figure 2, which reports the accumulated values (assuming reinvestment of dividends) of an investment of one dollar in January of 1926 in two portfolios: the S&P500 and a portfolio of small-firm stocks. While small firms suffered more in the Great Depression, their growth since that episode has been far more dramatic than the growth in a portfolio represented by the S&P500.

According to the EMH, the small-firm effect should be due solely to higher beta coefficients for small stocks; in other words, the higher rate of return is solely due to higher risks. Any unwarranted growth should lead investors to restructure their portfolios to include more small-cap firms, thereby driving the price of small-cap stocks up relative to high-capitalization stocks and restoring a "normal" relationship in which all firms enjoy the same rate of return, after adjustment for risk. The evidence suggests, however, that the higher return on small-cap stocks is not attributable to higher risk. While in recent years the small-firm effect has disappeared, the puzzle is that it existed for so many years, in spite of general awareness that it was there.

The Closed-End Mutual Fund Puzzle

Another well-known anomaly involving a specific class of firms is the Closed-End Mutual Fund Puzzle, reflected in the discounts (and occasional premia) on closed-end mutual funds (Lee, Shleifer and Thaler 1990b; Malkiel 1977). Closed-end mutual funds differ from open-end mutual funds in that open-end funds keep the prices of their shares at the net asset value (NAV) by promising to buy or sell any amount of their shares at NAV. Closed-end funds, on the other hand, issue a fixed number of shares at inception, and any trading in those shares is between investors; this allows the closed-end fund share price to deviate from NAV, that is, closed-end funds can trade at either a discount or a premium. If the EMH is valid, then any sustained discount or premium on closed-end fund shares must be due to unique characteristics of the fund's assets or charter. In the absence of such distinguishing characteristics, any discounts or premiums would induce investors to engage in arbitrage that would eliminate the discount or premium. For example, an unwarranted discount would lead investors to buy the closed-end fund shares and sell short a portfolio of stocks identical to that held by the fund, thereby capturing a riskless increase in wealth equal to the discount. A premium, on the other hand, would induce investors to sell short the closed-end fund and buy an equivalent portfolio of stocks.

But closed-end fund shares typically sell at discounts, and the discounts are often substantial. Figure

Figure 2

Accumulated Value of $1 Invested in 1926

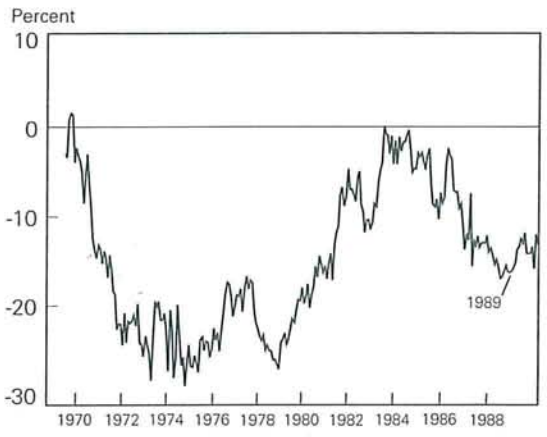
Source: Author's calculations, using Ibbotson (1990).

3 shows the average year-end discount in the period 1970–89 for seven major diversified closed-end fund companies.[5] It is clear that the discounts move inversely to stock prices; periods of bull markets, such as 1968–70 and 1982–86, are associated with low discounts, while bear markets (the 1970s and 1987) are associated with high discounts. Thus, the price paid for a dollar of closed-end fund assets is procyclical.

Several reasons are offered for closed-end fund discounts. First, because of potential capital gains taxes on unrealized appreciation, a new buyer of closed-end fund shares faces a tax liability if the fund should sell appreciated securities; this potential tax liability justifies paying a lower price than the market value of the underlying securities.[6] Second, closed-end funds might have limited asset marketability if they buy letter stock or privately placed debt, which cannot be sold to the public without incurring the expense of obtaining Securities and Exchange Commission (SEC) approval or the restrictions on corporate policy often required by public market investors. Third, agency costs, in the form of high management fees or lower management performance, might explain the discounts.

Figure 3

Average Premium (+) or Discount (-) on Seven Closed-Ended Funds



Note: The seven companies are listed in text footnote 5.
Data from Barron's.

Malkiel (1977) found that the discounts were larger than could be accounted for by these factors, and other work has confirmed that this appears to be a true anomaly. To this should be added another puzzle: at inception, the initial public offering (IPO) of closed-end fund shares must incur underwriting costs and, as a result, the shares must be priced at a premium over NAV, after which the price of seasoned shares typically moves to a discount within six months. Why would informed investors buy the IPO, thereby paying the underwriting costs via capital losses as discounts emerge? Clearly, something irrational is going on!

Weekend and January Effects

Another class of anomalies focuses on specific time periods or seasonalities. Cross (1973) reported evidence of a Weekend Effect, according to which weekends tend to be bad for stocks; large market decreases tend to occur between the close on Friday and the close on Monday. Later work showed that the weekend effect really occurs between the Friday close and the Monday opening. In Appendix 1 a weekend effect is added to the time series model of stock prices for the 2,713 trading days in the 1980s. The result is resounding statistical support for a weekend effect. A plausible explanation of the weekend effect is that firms and governments release good news during market trading, when it is readily absorbed, and store up bad news for after the close on Friday, when investors cannot react until the Monday opening.

In recent years the January Effect has received considerable attention; the rate of return on common stocks appears to be unusually high during January. The primary explanation is the existence of tax-loss selling at year end: investors sell their losing stocks before year end in order to obtain the tax savings from deducting those losses from capital gains realized during the year. The selling pressure in late December is then followed by buying pressure in January as investors return to desired portfolio compositions. However, this explanation is not consistent with the EMH, according to which investors with no capital gains taxes, such as pension funds, should identify any tendency toward abnormally low prices in December and should become buyers of stocks oversold in late December. This means that tax-loss selling should affect the ownership of shares but not their price.

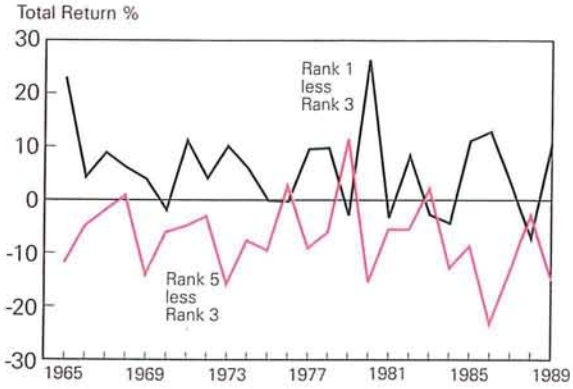
The January effect has been thoroughly investi-

gated, and has been found to be more complicated than originally thought. Keim (1983) has shown that the January effect appears to be due largely to price behavior in the first five trading days of January; it is really an Early-January Effect. Also, Reinganum (1983) found that the January effect and the small-firm effect are commingled: the January effect appears to exist primarily for small firms and, in fact, much of the small-firm effect occurs in January.

In the time-series analysis reported in Appendix 1, a test was added for the January effect and for an early-January effect. The results do not support a January effect of either type in the 1980s, at least for the S&P500. The fact that it does not appear for large firms, which dominate the S&P500 and are the firms of primary interest to institutional investors, is consistent with the EMH. Arbitrage by well-informed institutional investors appears to prevent any late-December selling pressure from affecting the share prices of large-capitalization firms. This does not provide any conclusions about the January effect for small firms.

Figure 4

Annual Excess Return on Stocks, Classified by Value Line Rank, 1965 to 1989



Source: Value Line Investment Survey.

The Value Line Enigma

Yet another well-known anomaly is the Value Line Enigma. The Value Line Investment Survey produces reports on 1700 publicly traded firms. As part of its service, Value Line ranks the common stocks of these firms in terms of their "timeliness," by which it means the desirability of purchasing the firm's shares. Value Line employs five timeliness ranks, from most timely (Rank 1) to least timely (Rank 5). Rank 3 is the designation for firms projected to increase in line with the market.

Figure 4 reports the annual average *excess* returns for Rank 1 and Rank 5 stocks. These returns are computed as the difference between the mean returns on stocks in the stated rank and the overall mean (Rank 3) returns. The computation assumes that the stocks are bought at the beginning of the year and sold at the end of the year, and transaction costs are not considered. It is clear that Rank 1 stocks generally perform better than average. In only five of the twenty-five years do Rank 1 stocks underperform the average; the probability of this happening by chance is only 0.00046.[7] Also, Rank 5 stocks tend to underperform. In only three of the twenty-five years do Rank 5 stocks perform better than average, and then the difference is small.

If the stock market is efficient, only one reason exists for higher rank stocks to generate higher returns: they have a higher level of market risk, that is, higher beta coefficients. Black (1971) found that the mean beta coefficients were roughly the same for stocks in each rank, concluding that the ranking system did have predictive value. However, Lee (1987) found that a stock's beta coefficient is inversely related to its Value Line rank: stocks for which purchase is timely tend to have higher betas. This suggests that the better performance of stocks ranked 1 and 2 is, at least in part, due to the higher average returns normally associated with higher risk.

Holloway (1981) examined the value of both active and passive trading strategies based on the Value Line Ranking System. An active trading policy was defined as purchasing Rank 1 stocks at the beginning of a year and holding them until the earliest of either the end of the year or a downgrade of the stock's Value Line rank, at which time the stock would be replaced by another Rank 1 stock to be held until year end. A passive, or buy-and-hold, strategy was defined as purchasing Rank 1 stocks at the outset of a year and selling them at year end. The active trading strategy generated higher returns than did the passive strategy when transaction costs were not considered, but was inferior to the buy-and-hold strategy when reasonable transaction costs were as-

sessed. Hence, *active* trading using the Value Line ranking system is not a profitable strategy for investors.

However, Holloway found that even after adjustments for transactions costs and for risk, a passive strategy using Rank 1 stocks outperformed a passive strategy using Rank 3 stocks; the Value Line Ranking System did provide profitable information for those who are willing to buy and hold. It is noteworthy that this advantage existed even when adjustments were made for both transaction costs and risk (beta).

III. "Modern" Evidence of Inefficiency

The previous section reports the results of "traditional" approaches to assessing the EMH: examination of specific examples of departures from the EMH, called anomalies. During the 1980s several "modern" approaches were developed. These are the topic of this section.

Excess Volatility of Stock Prices

One of the more controversial "modern" tests of the EMH is based on the observed volatility of stock prices. Leroy and Porter (1981) and Shiller (1981) concluded that the observed amount of stock price volatility is too great to be consistent with the EMH. In order to understand this "excess volatility" argument, refer back to the first definition of an efficient market: a market is efficient if the price of the asset is an optimal forecast of the fundamental value, that is, if $P_t = E(P_t^* | \Omega_t)$.

The logic of the excess volatility argument is based upon a property of statistical theory: the optimal forecast of a random variable must, on average, vary by no more than the amount of variation in the random variable being forecasted. Thus, if the market price is an optimal forecast of the fundamental value—as the EMH implies—it should vary less than (and certainly no more than) the fundamental value.

A formal statement of the excess volatility argument is that the relationship between the fundamental price under the actual state (s) and the optimal forecast of the fundamental price is

$$(4) \qquad P_s^* = E(P^* | \Omega_t) + \epsilon_s$$

where ϵ_s is a random variable that measures the deviation between the fundamental value for the state which actually occurs, P_s^*, and the optimal forecast of the fundamental value, $E(P^* | \Omega_t)$. If the forecast is optimal, these deviations must be random and uncorrelated with the forecast itself. Now, the EMH implies that $P = E(P_s^* | \Omega_t)$, which means

$$(4') \qquad P_s^* = P + \epsilon_s.$$

In other words, the correct price (conditional on knowing the true state) is equal to the market price plus a random term, denoted by ϵ_s, which measures the surprise resulting when the true state is known. This random term must be uncorrelated with P, because P is the optimal forecast and, therefore, already reflects any systematic information.

This provides the basis for the variance bounds tests of the EMH. Equation (4') shows that the variance of the fundamental price is equal to the variance of the market price *plus* the variance of the surprise. Turning this around produces the following relationship:

$$(4'') \qquad VAR(P_t) = VAR(P^*_t) - VAR(\epsilon_t).$$

Because variances *must* be non-negative, if the EMH is valid the variance of the market price must be no greater than the variance of the fundamental value, or:

$$(4''') \qquad VAR(P_t) \leq VAR(P_t^*).$$

Consider the following simple example, summarized in Table 1. Assume three states of the world, in each of which the dividend-price ratio is 10. In state 1 dividends paid at year end will be $10 and the

Table 1
Example of Variance Bounds Tests
Three States—Three Years

State (s)	Fundamental Value (P_t^*)	Probability of State s in Year 1	2	3
1	$100	.25	.50	.10
2	75	.50	.30	.80
3	40	.25	.20	.10
Market Price		$72.50	$ 80.50	$74.00
Modal Fundamental Value*		$75.00	$100.00	$75.00

*The modal fundamental value assumes that the most likely state occurs in each year.

fundamental value is $100, in state 2 dividends are $7.50 and the fundamental value is $75, and in state 3 dividends are $4 with a fundamental value of $40. These fundamental values are shown in column 2 of Table 1. At the beginning of each year the dividend to be paid at the end of the year is not known because the state that actually occurs is not known, but the *probability* of each state occurring is known. Therefore, at the beginning of each year investors know only the probability distribution of states and the dividend payment that each state entails. In this example, changes in the probability distribution across states correspond to the notion that new information is received by investors at the beginning of each year.

Under the null hypothesis of the EMH, the market price must vary by no more than the fundamental price. Any "excess" volatility is, therefore, a symptom of market inefficiency.

The "market's" problem is to determine a market price that best reflects that information. Table 1 assumes three years, with columns 3 to 5 showing the probability distribution of states in each year. The row marked "market price" shows the EMH market price, defined as the statistical expectation of fundamental values in each year. Thus, as time passes, the market price should increase from $72.50 to $80.50, then fall to $74.00; the sample standard deviation of the market price would be $4.25.

But the "correct" price, defined as the fundamental value associated with the realized state, would exhibit even larger movements. For example, if in each year the modal (most likely) state occurs, then the sequence of states is 2, 1, 2 and the fundamental values (row 5) would be $75 in year 1, $100 in year 2, and $75 in year 3. The sample standard deviation of these three "correct" prices would be $14.43, much greater than the sample standard deviation of the market price.[8] This result is consistent with the EMH.

Shiller's excess volatility tests were conducted as follows. He assumed a dividend valuation model in which the fundamental value is the present value of the perpetual stream of dividends resulting in each state-of-the-world. Using actual data on dividends paid over a very long period of time, and an assumption about the terminal price of shares, he calculated a time series for the fundamental value of the S&P500 index. He then compared the variance of that series with the variance of the observed values of the S&P500 and found that, if the discount rate was assumed to be constant, the variance of the market price was about six times the variance of the fundamental value—dramatic refutation of the EMH. However, if the discount rate was allowed to vary with interest rates (so that fundamental values exhibited greater variation), the market price had a variance about 1.5 times the variance of the fundamental price. In either case, the volatility of the stock market was greater than the upper bound implied by the EMH, leading Shiller to reject the EMH.

Under the null hypothesis of the EMH, the market price *must* vary by no more than the fundamental price. But Shiller's discovery implies either that the EMH is invalid or his test is invalid. This is a common problem of statistical tests: one must make assumptions about the world in order to construct any test, but one cannot know whether rejection of the null hypothesis is due to the invalidity of the hypothesis or to the invalidity of the assumptions.

The conclusion that excess volatility exists has been criticized for a number of reasons, each of which can be seen as a criticism of the test. Marsh and Merton (1986) disputed one of the assumptions underlying Shiller's test—that dividends are a stationary time series—and showed that if the process by which dividends are set is non-stationary, the EMH test is reversed: under the EMH, market prices *should* be more volatile than fundamental values. Kleidon (1986) has criticized the excess volatility test on statistical grounds, arguing that the Shiller test is an asymptotic test, assuming a very large sample of observations over time, and that the data available are necessarily finite, hence small-sample biases can weaken the test. In addition, the power of the test against reasonable alternative hypotheses is quite low, meaning that the test is not likely to reject the EMH when it should be rejected.

Whatever the validity of the excess volatility tests, they do provide an additional reason—other than observed anomalies—to doubt the validity of the EMH, and they have had a significant effect on the state of academic thinking about market efficiency.

Speculative Bubbles

It has long been a common practice to look back on dramatic collapses in asset prices and assign them to the bursting of a bubble. For example, following the October 1987 crash, many observers pointed out that stock prices had risen so rapidly in 1986 and 1987 that a bubble surely existed.

The notion of a "bubble" is a familiar one: a bubble reflects a difference between the fundamental value of an asset and its market price. Unfortunately, while the notion of a bubble has rhetorical force, it is a far more slippery concept than it appears. Clearly, a bubble is not merely a *random* deviation of price from value, for the law of large numbers suggests that purely random deviations will wash out over time without any necessity of collapse.

The bubble concept has been powerful because of the notion of self-fulfillment: bubbles are *self-fulfilling* departures of prices from fundamental values which continue until, for some reason, the conditions of self-fulfillment disappear. What do we mean by self-fulfilling bubbles? Recall that financial theory states that the market value of an asset (including dividends received) at the end of one period must be the market price at the end of the previous period, adjusted for growth at the required rate of return (r). That is, in equilibrium, where r is the required rate of return associated with the asset's risk level and E_t denotes an expectation conditional on information at time t:

$$(5) \qquad E_t(P_{t+1} + D_{t+1}) = (1 + r)P_t.$$

This difference equation, when solved recursively, gives the following stock price model, which is the well-known present discounted value model:

$$(6) \qquad P_t = \sum_{k=1}^{\infty} [(1 + r)^{-k} E_t D_{t+k}].$$

One definition of a self-fulfilling speculative bubble is that its presence does not violate this description of asset prices. In that case, called a "rational bubble," market observers could not see the presence of a bubble and would not behave in ways that eliminate it. But how would a bubble remain invisible? To do this, it must be true that its existence does not violate the process shown in equation (5). If we define B_t as the size of the bubble, we can see that if the bubble is expected to grow at the required rate of return, that is, if $E_t B_{t+1} = (1 + r)B_t$, the bubble will be viable. In this case investors do not care if they are paying for a bubble because they expect to get the required return on that investment.

This definition of a rational bubble implies some very strong restrictions on bubbles. One is that bubbles cannot be negative: in order to be self-fulfilling, a negative bubble must become more negative at the geometric rate r, but the stock price will grow at a rate less than r because dividends are paid.[9] From (5) we can see that $E_t(P_t + 1) = (1 + r)P_t - E_t(D_t + 1)$. Hence, a negative rational bubble must ultimately end in a zero price, a result that, once acknowledged, must lead to the elimination of the negative bubble. Thus, while the market price might be below the fundamental value at a specific point in time, it cannot be the result of a rational bubble.

Because there is no upward limit on prices, a positive bubble can exist, although with some implausible consequences. First, as time passes a positive rational bubble must represent an increasing proportion of the asset's price. This is because the bubble must grow at the rate r, while the price grows at a rate less than r because of dividend payments.

Bubbles are self-fulfilling departures of prices from fundamental values, which continue until, for some reason, the conditions of self-fulfillment disappear.

But the idea that investors can project an indefinite increase in the relative size of the bubble undermines the existence of the bubble. Surely, if investors understand that a positive bubble means that the bubble must be an increasingly important component of price, they will imagine that at some time the bubble must burst. But as soon as they realize that it *must* burst, it *will* burst!

For example, suppose investors believe that a positive bubble exists but that it will not burst until the year 2091. They must, then, realize that in the year 2090 the market price must reflect only the fundamental value because the year 2090 investors will not pay for a bubble knowing that it will disappear. But if the year 2090 price is the fundamental

value, no bubble can exist in 2089, and therefore the year 2089 price must be equal to the fundamental value. This chain of reasoning leads to the conclusion that a bubble cannot exist now. This will be true even if the collapse of a supposed bubble will not occur until after the passage of a very great (approaching infinite) time: as long as the resale price of the asset plays a negligibly small role in price determination, a bubble cannot exist![10]

If a rational bubble can never emerge, what is left of the notion of bubbles? Remember that a crucial assumption of the "bubbles cannot exist" paradigm is that investors behave as if they have an infinite time horizon. If investors have finite horizons, and plan to sell their shares before the present value of the sale becomes negligibly small, they will not project cash flows into the indefinite future, but will form judgments about the price at which the asset can be sold at the end of the horizon.[11] If, for example, the horizon is five years, the market price of the asset now will be described by the standard valuation equation $P_t = \Sigma_1^5[(1 + r)^{-k}E_tD_{t+k}] + (1+r)^{-5}E_tP_{t+5}$. If the expected resale price is simply the present value of expected dividends beyond that point, we are really back to the infinite-horizon model in which the ultimate resale price is irrelevant and bubbles cannot exist. For example, if $E_tP_{t+5} = \Sigma_6^\infty (1 + r)^{-k}E_tD_{t+k}$ we can see that the correct price will be described by equation (6).

Thus, the presence of a resale price whose expected value is not hinged to dividends beyond that point is necessary to the existence of rational bubbles. While it might be "rational" to use the infinite-horizon valuation model, it is not "realistic." Investors and traders do form judgments about the price at which they can sell assets, but they do not believe that the buyers are using an infinite-horizon model to decide the value of the asset.

Thus, rational bubbles are realistic descriptions of stock price performance; if the "market's" horizon is shorter than the time to the popping of a bubble, the bubble can continue. This is the essence of the "Greater Fool" explanation of speculative episodes: you will knowingly pay a price above fundamental value because you believe that someone later on will pay an even greater premium over fundamental value.

How should one go about testing the role of rational bubbles? This question is difficult to answer, for a rational bubble will not affect the sequence of prices until it breaks. The analysis of such low probability events is called the "peso problem": market prices will not reflect the effects of very low probability events even if they should have dramatic effects when they appear. Hence, it would be impossible to uncover a rational bubble as long as it exists. However, the disappearance of a bubble, such as a major decline in stock prices, can be examined to determine whether it was preceded by a speculative bubble in price.

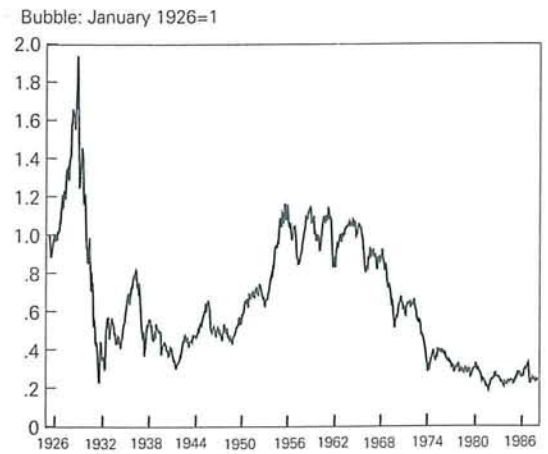
Using the Ibbotson (1990) data for monthly returns on common stocks (S&P500) and one-month Treasury bills, a measure of stock price bubbles for the period February 1926 to December 1988 was constructed. This was done by computing, for each month, the difference between the total return on common stocks and the required return. The required return was computed as the one-month Treasury bill rate plus a risk premium. Denoting the actual return as R_t and the required return as $(r_t + \theta)$, where r_t is the one-month Treasury bill rate and θ is the risk premium, the bubble at time t is:

$$B_t = [1 + R_t - (r_t + \theta)]B_{t-1}. \tag{7}$$

This approach assigns any difference between the observed return and the required return to

Figure 5

Estimated S&P 500 Stock Market Bubble



Source: Author's calculations, using Ibbotson (1990).

Table 2
Probability Model for Stock Market Crashes
Monthly Data 1926–88

	Size of Crash Over Next 12 Months					
	One Standard Deviation[a]		Two Standard Deviations		Three Standard Deviations	
CONSTANT	−1.5611	−1.3795	−2.4792	−1.2698	−3.7940	−1.9569
	(−10.81)	(−6.34)	(−12.38)	(−4.50)	(−10.49)	(−3.92)
$\log BUBBLE_{t-1}$	+.9327	+.8240	+2.1689	+1.3931	+1.7837	+.1897
	(+3.99)	(+3.24)	(+4.52)	(+2.99)	(+2.23)	(+.22)
TIME	n.a.	−.0007	n.a.	−.0058	n.a.	−.0139
		(−1.12)		(−4.36)		(−2.46)
Number of Months	755	755	755	755	755	755
Proportion Predicted	.8808	.8808	.9603	.9603	.9881	.9881
Mean Probability	.7019	.7024	.8611	.8754	.9411	.9532
Number of Months Followed by Crashes	90	90	30	30	9	9

Note: Numbers in parentheses are t-statistics. The parameters are estimated using a logit model, according to which Prob(crash) = 1/1{1 + exp[−(a + bX)]}, where X is the list of explanatory variables (logBUBBLE and TIME).
n.a. = not applicable.
[a] In the 755 months in the period 1926:2 to 1988:12, the change in log bubble over the next 12 months had a mean of −0.0208 and a standard deviation of 0.2226. Hence, a "one-standard-deviation" crash is defined as a 12-month change in the logarithm of the bubble by an amount of −0.2434 or less. A two-standard-deviation crash is a change of −0.4660 or less.

changes in a bubble. Assigning any arbitrary positive value to the initial bubble makes it possible to trace out the path of the bubble using this difference equation. Note that the value assigned to the initial bubble is irrelevant since our interest is in movements in a bubble, not in its absolute size.

Figure 5 shows the path of our measure of the bubble over the period February 1926 to December 1988, using the assumptions that the initial bubble is 1.0 in January of 1926 and that θ is the average risk premium over the entire sample period.[12] The time series shown in Figure 5 indicates a very large speculative bubble in the late 1920s as the return on stocks sharply exceeded the required return. This was followed by the crashes of 1929–32. The next bubble emerged in the 1955–65 period, when the bubble appeared to remain high for a considerable period of time before a prolonged "crash" lasting through the 1970s. The bubble fell to a low point in late 1982 that matched the lows of the 1930s.

The Crash of 1987 has often been attributed to a speculative bubble emerging as prices rose dramatically in 1986 and the first nine months of 1987. However, our bubble measure does not support this interpretation. While stock returns were above the required return, creating an expanding bubble, the size of the bubble in September 1987 was so small that it could not be used to predict the crash.

Does our measure of a stock market bubble have any predictive value? Clearly, the concept of a bubble is intended to explain market crashes, not simply mild and temporary price declines. Therefore, we have employed a probabilistic model to determine whether the *probability* of a future crash is related to the size of the bubble. In an efficient market no relationship exists but, if speculative bubbles do exist, the probability of a crash should be a direct function of the size of the bubble.

Table 2 reports our results for several different definitions of a crash. A "one standard deviation crash" occurs if over the *next* twelve months the change in (the logarithm of) the bubble is more than one standard deviation *below* the mean change.[13] We also consider crashes of two and three standard deviations. The probability of a crash is assumed to be described by a logistic function, and the result is a logit[14] model in which the probability of a crash over the next twelve months is a function of the logarithm of the bubble at the end of the previous month. In order to correct for any trends in the relationship, we have also added a linear trend variable (TIME).

The results for 1926–88 suggest that the bubble does have predictive value: for both one- and two-standard-deviation definitions of a crash, the variable

log $BUBBLE_{t-1}$ is statistically significant and has a positive sign: the bigger the bubble, the greater the probability of a crash in the next twelve months. The bubble size loses its predictive value if a crash is a three-standard deviation fall, but only 8 months fit this definition, and a model cannot predict events that occur so rarely.

Mean Reversion in Stock Returns

The phenomenon of mean reversion is the tendency for stocks that have enjoyed high (low) returns to exhibit lower (higher) returns in the future; that is, returns appear to regress toward the mean. A seminal test of mean reversion was undertaken by Fama and French (1988), who regressed the rate of return for a holding period of N months upon the rate of return during the previous N months. For example, they regressed the return over an 18-month period upon the return over the previous 18 months; a negative slope coefficient indicates mean reversion over an 18-month horizon. Fama and French found evidence of mean reversion for holding periods longer than 18 months.

While we have some skepticism about the Fama-French tests (the results appeared to be due primarily to the inclusion of the 1930s in the sample period), the phenomenon of mean reversion has been supported by other tests. Poterba and Summers (1988) found that the variances of holding-period returns do not increase in proportion to the length of the holding period, an indication of mean reversion.[15] Other variants of mean reversion tests also confirm the existence of mean reversion. For example, De Bondt and Thaler (1985) have found evidence of both a winner's curse and a loser's blessing in stock prices. Stocks that have experienced a recent reduction in their P/E ratios tend to have higher rates of return than equivalent stocks that have not been "losers," while stocks that have experienced increases in their P/Es tend to be losers in subsequent periods. The loser's blessing appears to be more dramatic than the winner's curse.[16]

Mean reversion can be thought of as another anomaly, but it is really much more: it is the common theme in most tests of stock market efficiency.

Mean reversion can be thought of as another anomaly, but it is really much more: it is the common theme in most tests of stock market efficiency. For example, Shiller's excess volatility tests are an indirect test for mean reversion, and the analysis of bursting speculative bubbles is really an examination of sudden mean reversions.

IV. Why Is the Stock Market Inefficient?

Abundant evidence casts doubt on the Efficient Market Hypothesis. The natural next question is, "Why?" What aspects of investor behavior might account for these departures from the predictions of economic theory?

An important preliminary to answering this question is the observation that many of the anomalies shown in the previous section are really manifestations of one fundamental anomaly: the small-firm effect. For example, the January effect is primarily a characteristic of small firms (Keim 1983; Reinganum 1983), and the winner's curse and loser's blessing are also most prominent among small firms (De Bondt and Thaler 1985). Furthermore, the closed-end fund puzzle appears to be the result of the similarity of the markets for closed-end funds and for small-firm stock: Lee, Shleifer and Thaler (1990a) show that discounts on closed-end funds are highly correlated with performance of small-firm stocks, that institutions tend to shy away from both small-firm stocks and closed-end fund shares, and that the transaction size of both small-firm stocks and closed-end fund shares tends to be much lower than the transaction size for large-firm stocks.

This suggests a common denominator for many—but not all—of the departures from the EMH. They tend to be concentrated in stocks traded in relatively narrow markets where the "smart money" is not as likely to play. In short, inefficiencies might be associated with a form of market segmentation in which the EMH applies to stocks of large firms which are the province of financial institutions with access to research on fundamentals, while inefficiencies, and their associated profitable opportunities, appear to be concentrated among those who invest in the stocks of smaller firms, traded in less active markets with a lower quality of information.

Market Inefficiency and Market Segmentation

The claim that market participants can be segmented into highly informed and less informed investors, and that this fact is an important component of stock price determination, will create a sense of déjà vu: presenting it to market practitioners is a case of preaching to the choir. Indeed, the "smart money–dumb money" distinction has been around for as long as markets have existed, and it was enshrined in the work of early dissidents in the random walk debate. For example, Cootner (1964) argued that the profits of professional investors, who have low transactions costs, come from observing the random walk of stock prices produced by nonprofessionals and stepping in when prices wander sufficiently far from the efficient price.

Why has this view enjoyed a renaissance among academic economists? It is not merely because academics get to put notches on their guns when they disturb the conventional wisdom. Nor is it solely due to the more important reason that anomalies have become too numerous and well-documented to ignore. Each of these has played a role, but the fundamental reason is that only recently have economists provided a theoretical foundation for market segmentation.

EMH theorists rejected the market segmentation approach for several reasons. The first was that it clearly assumes irrational behavior, with the unsophisticated investors (henceforth called small investors) somehow driving prices of stocks (primarily of small firms) away from fundamental value. The second is that it assumes that the smart money allows this to happen and fails to step in when very small opportunities arise; in short, arbitrage is incomplete. The third problem is survival of the small investors; if small investors are buying high and selling low, as they must if they are giving large investors an opportunity for profits, then the population of small investors should diminish over time and the inefficiencies should disappear.

The possibility that some investors are "irrational" is not sufficient to induce inefficiency; nobody believes that all investors are rational, and so long as these investors are infra-marginal they are merely giving profits to large investors. True, theorists do not like irrationality, and might question why it should exist; but that is a question of psychology, not of economics. Furthermore, the proposition that irrationality is self-correcting because the irrational players incur losses and leave the game does not work well for two reasons. First, the mortality of investors, and the difficulty of transmitting wisdom and experience to the young, mean that a new crop of investors is always emerging which, if given sufficient endowments, can become players. The 1980s were an example of this, with young professionals having limited experience handling large amounts of money. Second, as we shall argue, it is possible that irrational investors do, in fact, get rewarded and are not eliminated.

It is possible that irrational investors do, in fact, get rewarded and are not eliminated.

The fundamental objection to market segmentation is the arbitrage objection. Large investors (it is argued) will ensure that prices must be nearly efficient. Because they care only about intrinsic value, if prices diverge from the efficient price, they will engage in arbitrage to restore the equality. The arbitrage objection can be dealt with by forgoing two implicit assumptions: that investors never plan to liquidate their stock positions, and that riskless arbitrage is possible. Each of these assumptions is addressed in turn.

As in the discussion of speculative bubbles, if investors have finite horizons, they will be concerned about the resale price of the security and form judgments about what that will be at the end of their horizon. One way of forming those judgments—the way proposed by traditional finance theory—is to estimate the future price as a present value of dividends received from that point on; that simply brings in the infinite life assumption through the back door. Another approach—which seems more plausible—is to recognize that investors are concerned with resale price and that their forecasts of resale price may well not reflect solely their judgments about future dividends; perhaps even more important will be their estimates of what other investors will be willing to pay. This was stated clearly by John Maynard Keynes (1964, pp. 155–56), who said of professional investors and speculators:

> They are concerned, not with what an investment is really worth to a man who buys it "for keeps", but with

> what the market will value it at, under the influence of mass psychology, three months or a year hence . . . professional investment may be likened to those newspaper competitions in which the competitors have to pick out the six prettiest faces from a hundred photographs, the prize being awarded to the competitor whose choice most nearly corresponds to the average preferences of the competitors as a whole . . . We have reached the degree where we devote our intelligences to anticipating what average opinion expects average opinion to be.

But (an efficient market adherent might respond) even if many investors are forming judgments about resale prices that differ from fundamental value, a well-financed body of highly informed investors can prevent that from affecting market prices by engaging in riskless arbitrage. The response to this is that few opportunities for riskless arbitrage exist, and to the extent that arbitrage involves some risk, risk-averse investors will require a positive expected return (above opportunity costs), allowing inefficient pricing to continue.

An example of this is discounts on closed-end funds. Clearly, the market for closed-end fund shares must be dominated by investors who adopt a high probability that resale prices will deviate from fundamental values. If this were not true, investors would estimate that resale prices would be equal to future fundamental value, and riskless arbitrage would ensure that the current price reflects current fundamental value. For example, if current investors think that the future resale price will be above fundamental value, they will buy the closed-end fund shares and sell short a bundle of shares that replicate the closed-end fund portfolio. By doing this, they will enjoy profits, and their profit-seeking activities will ensure that current price is equal to current fundamental value.

Noise Trading

We have argued that prices *do* diverge systematically from fundamental values because prices *can* diverge systematically from fundamental values, because even well-informed investors are risk averse and will not engage in sufficient arbitrage activity to prevent this. This view has been formalized recently by a model of irrationality called "noise trading" by its proponents (Black 1986; Shleifer and Summers 1990). The noise trading model proposes that an important segment of the market consists of investors who bid prices away from fundamentals, thus introducing "noise" into stock prices. This noise, or "investor sentiment," is sufficiently broad in its impact, affecting many stocks, that investors cannot avoid it by diversification and must accept it as a source of systematic risk. Because it is systematic and undiversifiable, the noise affects the rate of return investors require on stocks and, therefore, market prices. Not all stocks are affected equally by noise risk—stocks of well-established firms, which are traded among informed investors, might not carry much noise risk, while stocks of small firms are more likely to bear this risk.

A simple model of noise trading is presented by DeLong, Shleifer, Summers and Waldmann (1990). This model, discussed in more detail in Appendix 2, assumes that young investors buy stocks and old investors sell stocks (to the young) to live on in their dotage. Sophisticated investors form optimal forecasts of the future price, but unsophisticated investors, called "noise traders," develop biased forecasts. Because sophisticated investors are risk averse and arbitrage is risky due to the possibility that the extent of price misperception by noise traders might change, sophisticated investors will not fully arbitrage away the influence of noise trading. Thus, noise traders can drive the market price away from the fundamental value.

The noise trading model proposes that an important segment of the market consists of investors who bid prices away from fundamentals, thus introducing "noise" into stock prices.

In the DeLong-Shleifer-Summers-Waldmann model, the degree of price misperception exhibited by noise traders—the difference between their forecasts and optimal forecasts—is assumed to be a random variable (denoted as ρ), which follows a normal probability distribution with mean ρ^* and variance σ^2. If $\rho^* = 0$, noise traders agree with sophisticated traders "on average," but noise trader forecasts will temporarily differ from sophisticated forecasts at any moment. In this case the equilibrium

price of an asset will normally be below the fundamental value, the discount being necessary to compensate sophisticated traders for the risk that stock prices will deviate from fundamental value even more in the future. If noise traders are pessimistic ($\rho^* < 0$), the normal discount from fundamental value will be higher, while if noise traders are optimistic ($\rho^* > 0$), the normal discount will be lower or a premium might emerge. In addition to the normal discount arising from the average price misperception of noise traders, there is a temporary random discount due to temporary variations in optimism and pessimism ($\rho - \rho^*$).

The EMH adherent would ask why sophisticated investors do not dominate the market and, through arbitrage, force the market price to equal the fundamental value. If this did occur, the perceptions of noise traders would alter the ownership of stocks (sophisticated traders holding more when noise traders are pessimistic and less when noise traders are optimistic), but price misperceptions would not affect the market price of an asset.

The answer is, as noted above, that arbitrage is not riskless: no individual sophisticated trader can know that all sophisticated traders together will force equality of the market price with the fundamental value. There is always the possibility that noise traders will influence stock prices and that the sophisticated trader, when he arrives at the end of his horizon, will be forced to sell at a price even further below fundamental value than his cost.

This model is overly simple, designed for expository purposes and not as a strict representation of reality. But it does explain a number of important phenomena. For example, it explains the excess volatility of common stock prices found by Shiller: in the absence of noise trading, stock prices would always equal fundamental values, but with noise trading, stock prices will be more volatile than fundamental values because of the changing perceptions of noise traders. It also can explain the small-firm effect, and the related anomalies (for example, the January effect, the loser's blessing, the closed-end fund puzzle): the small-firm effect exists because the noise risk is higher among small firms, which are not as favored by sophisticated traders and in which noise traders play a larger role.

Furthermore, this simple model explains how the phenomenon of noise trading can persist. Friedman (1953) argued that, in the long run, prices must conform to fundamentals because speculators who paid incorrect prices would either go broke if they tended to buy high and sell low, or would force prices to equal fundamental value if they were sharp enough to buy low and sell high. In either case, what we now call noise trading would be a temporary phenomenon. In this model, however, noise traders create a more risky environment, but because their effects are pervasive and not idiosyncratic to individual stocks, the risk is not diversifiable and must earn a reward. Indeed, not only is noise trading consistent with an average return above the riskless rate, but it can be consistent with a higher average return for noise traders than for sophisticated investors if, as seems likely, noise traders tend to invest more heavily in noise-laden stocks, hence earning more of the risk premium associated with small stocks.

Noise Trading with Fads: A Simulation

The noise trading model can easily generate models of stock price movements that mimic the sharp breaks and apparent patterns visible in the real data. The crux of this is the possibility of "fads" that affect investor sentiment, measured by ρ. Small changes in these opinions can translate into very large changes in stock prices. For example, using equation (A2.1) in Appendix 2, we can calculate the effect of a change in investor sentiment of noise traders.

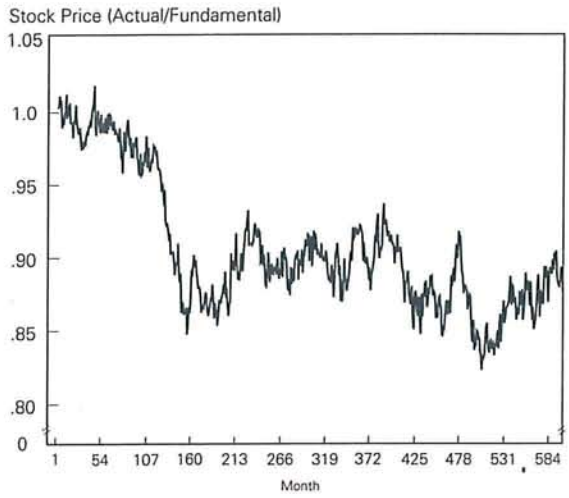
As noted above, noise trader price perceptions can be decomposed into two types: the normal perception is the average value of ρ, denoted by ρ^*, and is the degree of optimism that prevails over fairly lengthy periods; temporary price perceptions, denoted by $\rho - \rho^*$, prevail at any moment of time but do not affect the trend of stock prices.

Changes in the normal price perception, or ρ^*, can result in very large changes in stock prices. According to equation (A2.1), a change in ρ^* induces a change in stock price by the amount $(\mu/r)\ \rho^*$, where μ is the proportion of traders who are noise traders and r is the real interest rate. Assuming $\mu = 0.05$ and $r = 0.0042$ per month (5 percent per year), we calculate $(\mu/r) = 11.90$: a change in ρ^* by 0.01 (or 1 percent of fundamental value) will alter the stock price by 0.1190, or about 12 per cent of fundamental value.

Changes in stock prices due to variations in investor sentiment will be random so long as the normal price perception of noise traders is constant. However, if ρ^* is serially correlated—its current value depends on previous values—stock price changes can exhibit sharp breaks that do not conform to a model

Figure 6

Simulated 50-Year Stock Price History



Source: See Appendix 2.

of simple random variation about an equilibrium level. This is likely to happen when there are fads in perceptions, as when optimism is reinforced by earlier phases of optimism.

To illustrate this, we have simulated monthly behavior of a hypothetical stock market using equation (A2.1). The settings for the crucial parameters are discussed in Appendix 2. In order to introduce the possibility of realistic results from a noise trading model we need to add the notion of "fads and fashions," in which investor interactions create waves of investor sentiment. Our way of introducing the possibility of contagious behavior is to assume that "opinion" follows a random walk shown by the autoregressive process:

$$\rho_t^* = \rho_{t-1}^* + \epsilon_t \tag{8}$$

where ϵ_t is white noise with mean $E(\epsilon_t) = 0$ and constant variance σ_ϵ^2. In this simple case, investors standing at a moment of time will forecast a constant value of ρ^* because zero is their optimal forecast of ϵ for every period. But the actual value of ρ^* will follow a path determined by equation (8). Because stock prices are very sensitive to ρ^*, this can result in realistic stock market cycles.

The results of one such experiment are shown in Figure 6. The results show what appear to be systematic patterns in the stock price imposed upon a bear market, in which the stock price falls from its fundamental value of 1.0 to about 0.88 at the end of the 50 years (600 months). Recall that the results are stock prices relative to fundamental value, so the figure does not mean that stock prices fall to 88 percent of the original value; an upward trend in the fundamental value could allow the stock price to rise even though, because of noise trading, it is not rising as fast as it should.

Repeated experiments will show essentially the same patterns of prolonged departures from fundamental value, punctuated by sharp breaks in price, though the sequence of bull and bear markets will be different in each simulation. It is clear that even with mild fads in noise trader misperceptions, there can be dramatic and apparently systematic cycles in stock prices. While this does not prove that noise trading is the source of the kind of stock market cycles we observe, it does show that noise trading, supplemented with contagious investor interactions or fads, is a plausible way of explaining observed stock price behavior.

V. Summary and Conclusions

This paper assesses the current state of the efficient market hypothesis, which was the conventional wisdom among academic economists in the 1970s and most of the 1980s. It reviews the empirical evidence and concludes that it provides an overwhelming case against the efficient market hypothesis. This evidence exists in the form of a number of well-established anomalies—the small firm effect, the closed-end fund puzzle, the Value Line enigma, the loser's blessing and winner's curse, and a variety of anomalies surrounding seasonality, such as the January effect and the weekend effect. Many of these anomalies are more pronounced among small-firm stocks, suggesting that the efficient market hypothesis might be more appropriate for stock of large firms, but analysis of the S&P500, which is dominated by large firms, also finds important anomalies such as a weekend effect, slow mean reversions in returns, and stock price volatility in excess of the amount predicted by fundamentals.

These anomalies can be explained by resorting to a model of "noise trading," in which markets are segmented with the "smart money" enforcing a high

degree of efficiency in the pricing of stocks of large firms while less informed traders dominate the market for small firms. This model can explain many of the anomalies, and it can generate cycles in stock prices that are very similar to those observed in the real world.

Our fundamental conclusion is that the efficient markets hypothesis is having a near-death experience and is very likely to succumb unless new technology, as yet unknown, can revive it. This conclusion has a number of policy implications. The fundamental implication is that security market inefficiency provides an economic foundation for public policy interventions in security markets. Clearly, if markets are efficient, hence conforming to the paradigm of pure competition, there is little reason for a security market policy: the market works to correct imbalances and to efficiently disseminate information.

Our fundamental conclusion is that the efficient market hypothesis is having a near-death experience and is very likely to succumb unless new technology, as yet unknown, can revive it.

However, if inefficiencies do abound, reflecting barriers to entry in transactions or inefficient collection, processing, and dissemination of information, there might be a role for public policy. For example, the existence of the Securities and Exchange Commission, whose primary function is to ensure equal access to relevant information, would be questionable in a world with an efficient market for information; in that case market prices would more accurately reflect all relevant information. As another example, in a world of efficient markets, sharp changes in prices, such as the October 1987 break, would reflect dramatic changes in fundamentals and should not elicit public policy responses. But in an inefficient market, in which investor sentiment clouds the influence of fundamentals, policies designed to mitigate price changes (daily price limits, market closings under certain conditions) might be appropriate.

The objective of this paper is not an examination of sharp and maintained price breaks such as October of 1987. But the noise trading model does suggest one reason why that break appeared to be lasting in its effect on stock prices. To the extent that the price break was not associated with changes in fundamentals (and it is widely agreed that it was not), it could have adversely affected investor sentiment, inducing prices to remain below fundamental values for prolonged periods. A useful analogy is the prevalence of discounts on closed-end mutual funds—these discounts appear to exist because investors recognize that they might get larger.

The purpose of the third paper in this trilogy will be to pursue the question of dramatic breaks in stock prices, and to investigate the wisdom and efficacy of policies designed to address potential problems of short-run stock price volatility that arise from stock market inefficiency.

Appendix 1: Time Series Analysis of Daily Stock Prices in the 1980s

In this appendix we report some results of tests for random walks and for specific anomalies using daily stock price data for the 1980s. Figure A-1 shows the data: the daily closing price of the S&P500 for each of the 2,713 trading days from January 2, 1980 to September 21, 1990. The chart reports the logarithm of the price index, rather than the index itself, for two reasons. First, analysis of stock prices is usually done on the logarithm of the price because desirable statistical properties are the result; in particular, the changes in log stock prices are close to normally distributed, a property that allows a broad range of statistical tools to be brought to bear. Second, a graph of the log price has the property that the slope of the line measures the percentage rate of change of the price; for example, the chart shows that the rate of increase in stock prices was particularly high from mid-1982 through mid-1983, then slowed somewhat until the October 19, 1987 break, and after that break, the rate of increase was lower than it had been in the earlier bull market periods.

Are Daily Stock Prices a Random Walk?

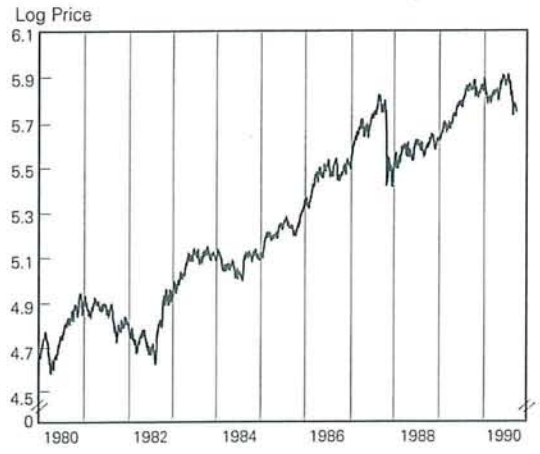
An attempt to fit this model to the daily S&P500 closing prices for 1980–90 failed to support the random walk hypothesis: while the intercept and slope coefficients were consistent with the EMH, the residuals were not white noise, but showed significant autocorrelation. Further experimentation using time series methods led us to conclude that the log of the daily closing value of the S&P500 corresponded to an Integrated Moving Average (IMA) model; we found that an IMA(1,5) model, using first differences and five daily moving average terms, was sufficient to eliminate autocorrelation. This equation, in which movements in the first difference of logP are described by a five-day moving average of white noise terms, is reported as equation 1 in Table A-1.

The moving average coefficients are statistically significant, and reveal the following pattern. If there is a downward shock ("crash") in the price change, the following adjustments will occur: on the following day the price change will be slightly more than the normal amount, after which it will increase at a slightly below-normal rate for three days, ending with a slightly above-normal increase on the fifth day. After five days, if no other shocks have occurred, the abnormal behavior is over. Given the short period of fluctuation, it is no surprise that longer-term data (such as monthly data) do not reveal departures from the EMH.

Thus, our data lead us to reject the random walk implications of the EMH for such short intervals of time as one day. However, the departure from a random walk is not of major economic significance; in short, it is not "bankable." First, the low coefficient of determination ($\overline{R}^2 = 0.01$) tells us that while the moving average terms are statistically significant, they explain only about 1 percent of the variation in the change in log price—there is a high probability that any potential profits from trading strategies based on the knowledge of the time series structure will be swamped by random variations. Second, even at its most profitable, the optimal trading strategy might not cover its costs: the optimal strategy would be to buy (sell) the day *after* a major fall (rise) in prices, then sell (buy back) the portfolio after it is held for four days. If we calculate the profits from doing this after a major price decline (defined as a decline greater than all but only 10 percent of price declines) the profits are only about 1.4 percent of the initial cost; this would not cover retail transactions costs, though it could cover institutional transaction costs.

Figure A-1

Logarithm of Daily S&P 500 Closing Price



Source: Data Resources, Inc.

Is There a Weekend Effect in the 1980s?

In order to examine the Weekend Effect during the 1980s we have re-estimated equation 1 of Table A-1 by adding two dummy variables: WKEND, which has a value of 1 if the trading day is a Monday, zero otherwise, and HOLIDAY, which has a value of 1 if the current trading day was preceded by a one-day holiday. The five-day moving average behavior reported in Table A-1, equation 1, is reproduced in equation 2. In addition, the WKEND dummy variable has a coefficient that is both negative and statistically significant. Thus, the daily data for the 1980s do contain a significant Weekend Effect. The HOLIDAY dummy is not statistically significant, indicating that one-day closings are not associated with systematic differences in price behavior.[17]

The January Effect in the 1980s

Equations 3 and 4 of Table A-1 incorporate dummy variables for the January Effect. Equation 3 includes JAN-

Table A-1
Tests of Random Walk Hypothesis
IMA(1,5) Model, Dependent Variable = Δ logP

Independent Variable	Equation 1	2	3	4
Constant	+.0004 (1.85)	+.0007 (3.12)	+.0007 (2.80)	+.0007 (3.07)
MA(1)	+.0526 (2.74)	+.0546 (2.84)	+.0546 (2.84)	+.0546 (2.84)
MA(2)	−.0370 (1.92)	−.0369 (1.92)	−.0369 (1.92)	−.0369 (1.92)
MA(3)	−.0200 (1.04)	−.0193 (1.00)	−.0192 (.99)	−.0193 (1.00)
MA(4)	−.0548 (2.85)	−.0544 (2.83)	−.0541 (2.81)	−.0544 (2.83)
MA(5)	+.0575 (2.85)	+.0518 (2.69)	+.0518 (2.69)	+.0517 (2.69)
WKEND	n.a.	−.0017 (3.19)	−.0017 (3.19)	−.0017 (3.19)
HOLIDAY	n.a.	+.0007 (.23)	+.0006 (.19)	+.0006 (.21)
JANUARY	n.a.	n.a.	+.0006 (.82)	n.a.
EARLYJAN	n.a.	n.a.	n.a.	+.0007 (.19)
$\bar{R}^2$	.0089	.0120	.0119	.0117
SEE	.0109	.0109	.0108	.0108
Q(156)	169.7670	169.3580	169.1590	169.2850
p	.2179	.2248	.2281	.2260

Note: The sample period is January 2, 1980–September 21, 1990. Numbers in parentheses are absolute values of t-statistics. p is the probability level of the Q-statistic (156df).
n.a. = not applicable

UARY, defined as 1 for trading days in January and zero otherwise, and equation 4 includes EARLYJAN, a dummy variable defined as 1 in the first five trading days of January and zero otherwise. In neither case is the evidence consistent with a January Effect in the 1980s; neither coefficient is statistically significant. We cannot, however, conclude that the January Effect has disappeared; it is possible that it still exists for small firms, but that it does not exist for the large firms that are in the S&P500.

Our regressions do, however, allow us to conclude that the January Effect can no longer be used as a profitable strategy for a broad range of large firms listed on the major exchanges. The fact that it does not appear for large firms, which are the firms of primary interest to institutional investors, is consistent with the Efficient Markets Hypothesis: arbitrage by institutional investors prevents late-December selling pressure from affecting the share prices of large-capitalization firms, while small-cap stocks that have performed poorly do not have the attention of institutional investors and are oversold at year end.

Appendix 2: A Model of Noise Trading

The De Long-Shleifer-Summers-Waldmann model of noise trading results in a market price of common stock described by the following equation:

$$(A2.1) \quad P = 1 + \mu\rho^*/r + \mu(\rho - \rho^*)/(1 + r) - 2\gamma\mu^2\sigma^2/[r(1 + r)^2]$$

where μ is the proportion of investors who are noise traders, r is the real interest rate on riskless securities, and γ is the degree of absolute risk aversion, assumed to be the same for all investors. It is assumed that the degree of price misperception by noise traders, denoted by ρ, is normally distributed with mean ρ^* and variance σ^2. The price described by (A2.1) is the market price relative to the fundamental value, so P = 1.0 means that the price is equal to fundamental value.

The last three terms reflect the influence of noise trading; if $\mu = 0$, there are no noise traders and $P_t = 1$, that is, stock prices are determined by fundamentals alone. The second term reflects the influence of the mean amount of mispricing. If $\rho^* > 0$ ($\rho^* < 0$), noise traders are *normally* bullish (bearish), and to the extent that they are important in the market (measured by μ), this raises (reduces) stock prices; sophisticated traders will take opposing positions (such as reducing their stock holdings when prices are above fundamentals). The third term reflects the effect of *unusual* bullishness or bearishness of noise traders, measured by the random variable $(\rho - \rho^*)$; once again, the effect this has on stock prices depends on the relative numbers of noise traders.

Finally, the last term reflects the effect of uncertainty about the future degree of price misperception. The net effect of this is negative because the uncertainty imposed by noise traders will discourage investment by both sophisticated traders and noise traders, both of whom realize that price reversals can occur. This effect will be smaller, the less risk averse investors are (if $\gamma = 0$ nobody cares about risk so it will not affect prices) and the less important are noise traders (if $\mu = 0$ noise traders do not exist so they cannot affect prices). Given the values of γ and μ, the stock price will be negatively related to the size of the variance of noise trader misperceptions (σ^2).

If we measure the volatility of stock prices by the standard deviation conditional on information available in the previous period, for example, by $s = E_{t-1}(P_t - E_{t-1}P_t)^2$, we see that:

$$(A2.2) \quad s = \mu\sigma/(1 + r)$$

which says that volatility will be greater the larger the representation by noise traders, the larger the variability in their price misperceptions, and the lower the rate of interest.

If this were the end of the story, our noise trading model would predict that stock prices will deviate randomly around a normal value that is determined by the following equation:

(A2.3) $$P_t = 1 + \mu\rho^*/r - 2\gamma\mu^2\sigma^2/[r(1 + r)^2].$$

If ρ^* is zero, stocks will be chronically undervalued because all traders recognize that noise exists and that they might have to sell their assets at prices below their fundamental values. This "noise risk" cannot be eliminated by diversification, and results in a market price less than the fundamental value. Any deviations from this constant price level would be purely random, arising from temporary deviations of ρ from ρ^*. In short, the model would simulate volatility but not replicate the patterns of bull and bear markets we observe in the real world.

In order to introduce the possibility of realistic results from a noise trading model we need to add the notion of "fads and fashions," in which investor interactions create waves of investor sentiment. One way of introducing the possibility of contagious behavior is to assume that opinions follow a random walk shown by the autoregressive process:

(A2.4) $$\rho_t^* = \rho^*_{t-1} + \epsilon_t$$

where ϵ_t is white noise with mean $E(\epsilon_t) = 0$ and constant variance σ_ϵ^2. In this simple case, investors standing at a moment of time will forecast a constant value of ρ^* because zero is their optimal forecast of ϵ for every period. But the actual value of ρ^* will follow a path determined by equation (A2.4). Because stock prices are very sensitive to ρ^*, this can result in realistic stock market cycles.

In order to complete the simulation model, we assume $\mu = 0.05$, that is, noise traders represent 5 percent of the market. We also assume $\sigma_\epsilon = 0.005$, so that opinion (ρ^*) follows a simple random walk with the standard deviation of 0.005 in the steps.[18] Finally, we assume $\sigma = 0.01$, so that in 84 per cent of the trials (months) the random component in the degree of mispricing (for example, $\rho - \rho^*$) will be within ± 1 percent of the fundamental value of the stock.

Any movements in the simulated stock price must be due to either ϵ or $(\rho - \rho^*)$. We complete the simulation model by assuming that both ϵ and $(\rho - \rho^*)$ are normally distributed with zero means and the standard deviations assigned above. Using a random number generator to pick values of ϵ and $(\rho - \rho^*)$ for each of our "months," we can track the stock price. This experiment was done 600 times to simulate the stock price over a period of 50 years (600 months).

The results of one simulation are shown in Figure 6.

[1] One popular joke intended to discredit the EMH was about two economists walking along a street in Chicago (the bastion of the EMH): one observes a $20 bill lying on the sidewalk and begins to bend down to get it, but the other tells him not to bother, for if the bill were really there it would already have been picked up!

[2] Technical analysis is the use of historical information on stock prices to forecast future stock prices. Perhaps the best example of technical analysis is the Dow Theory, which identifies specific patterns in stock prices and uses them to form forecasts.

[3] For example, a simple dividend valuation model with a constant growth rate for dividends (g) and a constant real discount rate (r) implies $E(P_t^*) = D/(r - g)$; knowing current dividends per share (D) and using estimates of r and g allows one to test the EMH by regressing P_t on a variable defined as $D/(r - g)$ and any other variable in the information set Ω. Adjustment of the regression method for measurement error will, of course, be necessary.

[4] For expositional convenience we are ignoring the payment of dividends. If dividends are included, the valuation equation must be modified to read

$$E(P_{t+1}^* + D_{t+1}) = (1 + r)E(P_t^* | \Omega_t).$$

Thus, dividends are assumed to be reinvested.

[5] The companies are Adams Express (NYSE), Baker Fentress (OTC), General American Investors (NYSE), Lehman Corporation (NYSE), Source Capital (NYSE), Tri-Continental (NYSE), and Niagara Shares (NYSE).

[6] Of course, an open-end fund with unrealized appreciation exposes the investor to the same tax liability. But the liability cannot affect the open-end fund's share price because it is always equal to the net asset value; the only effect is to induce investors with high tax rates to prefer closed-end funds over open-end funds.

[7] If the Value Line ranking system is of no use, the probability that Rank 1 stocks will outperform Rank 3 stocks is 0.50. Under this null hypothesis the probability of 5 or fewer "failures" in 25 years is only 0.00046. This provides a strong reason to believe that the ranking system does have merit.

[8] The keen-eyed reader will observe that one can construct an example from Table 1 that shows the fundamental value varying less than the market price. For example, if the same state occurred in each year, the fundamental price would not vary at all! This is an artifact of the example: with a large number of possible states and a large number of time periods, the optimal forecast *must* vary by no more than the fundamental price.

[9] From (5) we can see that $E_t(P_{t+1}) = (1 + r)P_t - E_t(D_{t+1})$. If no dividends are expected, the price will be expected to grow at the rate r. If dividends are paid, the rate of increase in price will be less than r, but the rational bubble must grow at the rate r.

[10] Thus, the imposition of the transversality condition $\lim_{t\to\infty} [(1 + r)^{-t}P_t] = 0$ is sufficient to crush a nascent bubble.

[11] Hence a necessary condition for existence of a speculative bubble is a finite time horizon. This is not, however, sufficient. Tirole (1982) has shown that even with a finite time horizon, a speculative bubble cannot exist if expectations are rational, that is, if investors' forecasts are optimal. Hence, bubbles require both finite horizons and non-optimal forecasting. Stated differently, rational bubbles require inefficient markets.

[12] Attempts to identify time-variation in the risk premium were not successful, so the value of θ was set at the sample average for $(R_t - r_t)$; this was 0.0070 per month or 0.0838 per year.

[13] For the 755 months in the period 1926:2 to 1988:12, the change in log bubble over the next 12 months had a mean of −0.0208 and a standard deviation of 0.2226. Hence, a "one-standard-deviation crash" is defined as a 12-month change in the logarithm of the bubble by an amount of −0.2434 or less. A two-standard-deviation crash is a change of −0.4660 or less.

[14] The logit model assumes that the probability of an event (π) is a logistic function of the following form:
$\pi_t = 1/\{1 + \exp[-(\alpha + \beta xt)]\}$.

[15] The logic of the Poterba-Summers test is simple. Suppose that the one-period rate of return on stocks is approximated by the change in the logarithm of the price. Suppose further that—as many financial studies assume—the change in the logarithm can be represented by a constant plus a random error term, so $\log P_{t+1} - \log P_t = \mu + \epsilon_{t+1}$. Then the average return over N periods is approximately $\log P_{t+N} - \log P_t = N\mu + (\epsilon_{t+1} + \epsilon_{t+2} \ldots + \epsilon_{t+N})$.

If the EMH is correct, the ϵ's are identically and independently distributed. Denoting the variance of ϵ as σ^2, the variance of the N period return is $VAR(\log P_{t+N} - \log P_t) = N\sigma^2$: the variance of returns is proportional to the period over which the returns are experienced. If, as Summers and Poterba conclude, the variance increases less than in proportion to the period, the return on stocks is mean-reverting.

[16] The term "winner's curse" is used here in a different way than it is used in discussing the effects of mergers and acquisitions. In that context, the winner's curse is the tendency of those who outbid others to pay too high a price for the acquired firm.

[17] We also found that three-day weekends were no different from two-day weekends; the extra day makes no difference, just as a one-day holiday makes no difference.

[18] The other parameter in the model (γ) really plays no important role in the simulation. We set it at $\gamma = 0.10$.

References

Black, Fischer. 1971. "Yes, Virginia, There Is Hope: Tests of the Value Line Ranking System." Center for Research in Security Prices, Graduate School of Business, University of Chicago, May.

———. 1986. "Noise." *Journal of Finance*, July, pp. 529–543.

Cootner, Paul H. 1964. "Stock Prices: Random vs. Systematic Changes." In Paul H. Cootner, ed. *The Random Character of Stock Market Prices*. Cambridge, MA: The MIT Press.

Cross, Frank. 1973. "The Behavior of Stock Prices on Fridays and Mondays." *Financial Analysts Journal*, November/December, pp. 67–69.

De Bondt, Werner and Richard H. Thaler. 1985. "Does the Stock Market Overreact?" *Journal of Finance*, July, pp. 793–805.

De Long, J. Bradford, Andrei Shleifer, Lawrence H. Summers, and Robert J.Waldmann. 1990. "Noise Trader Risk in Financial Markets." *Journal of Political Economy*, August, pp. 703–738.

Fama, Eugene F. and Kenneth R. French. 1988. "Permanent and Temporary Components of Stock Prices." *Journal of Political Economy*, April, pp. 246–273.

Fortune, Peter. 1989. "An Assessment of Financial Market Volatility: Bills, Bonds and Stocks." *New England Economic Review*, November/December, pp. 13–28.

Friedman, Milton. 1953. "The Case for Flexible Exchange Rates." In Friedman, M., *Essays in Positive Economics*, pp. 157–302. Chicago, IL: University of Chicago Press.

Holloway, Clark. 1981. "A Note on Testing an Aggressive Investment Strategy Using Value Line Ranks." *Journal of Finance*, June, pp. 711–719.

Ibbotson Associates. 1990. *Stocks, Bonds, Bills and Inflation: 1990 Yearbook*. Chicago, IL: Ibbotson Associates.

Keim, Donald. 1983. "Size-Related Anomalies and Stock Return Seasonality." *Journal of Financial Economics*, June, pp. 13–22.

———. 1986. "The CAPM and Equity Return Irregularities." *Financial Analysts Journal*, May/June, pp. 19–34.

Keynes, John Maynard. 1964. *The General Theory of Employment, Interest and Money*. New York, NY: Harcourt, Brace and World.

Kleidon, Allan W. 1986. "Variance Bounds Tests and Stock Price Valuation Models." *Journal of Political Economy*, October, pp. 953–1001.

Lee, Cheng F. 1987. "Value Line Investment Survey Rank Changes and Beta Coefficients." *Financial Analysts Journal*, September/October, pp. 70–72.

Lee, Charles, Andrei Shleifer and Richard Thaler. 1990a. "Investor Sentiment and the Closed-End Fund Puzzle." Johnson Graduate School of Business, Cornell University, January.

———. 1990b. "Closed-End Mutual Funds." *Journal of Economic Perspectives*, Fall, pp. 153–164.

LeRoy, Stephen F. and Richard D. Porter. 1981. "The Present-Value Relation: Tests Based on Implied Variance Bounds." *Econometrica*, May, pp. 555–574.

Lucas, Robert E. 1978. "Asset Prices in an Exchange Economy." *Econometrica*, November, pp. 1429–1445.

Malkiel, Burton. 1977. "The Valuation of Closed-End Investment Company Shares." *Journal of Finance*, June, pp. 847–859.

Markowitz, Harry M. 1959. *Portfolio Selection: Efficient Diversification of Investments*. New York, NY: John Wiley and Sons.

Marsh, Terry A. and Robert C. Merton. 1986. "Dividend Variability and Variance Bounds Tests for the Rationality of Stock Prices." *The American Economic Review*, June, pp. 483–498.

Poterba, James M. and Lawrence H. Summers. 1988. "Mean Reversion in Stock Prices: Evidence and Implications." *Journal of Financial Economics*, February, pp. 27–59.

Reinganum, Marc. 1983. "The Anomalous Stock Market Behavior of Small Firms in January: Empirical Tests for Tax-Selling Effects." *Journal of Financial Economics*, June, pp. 89–104.

Samuelson, Paul. 1965. "Proof That Properly Anticipated Prices Fluctuate Randomly." *Industrial Management Review*, Spring, pp. 41–49.

Schwert, G. William. 1989. "Why Does Stock Market Volatility Change Over Time?" *Journal of Finance*, December, pp. 1115–1153.

Sharpe, William. 1964. "Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk." *Journal of Finance*, September, pp. 425–442.

Shleifer, Andrei and Lawrence H. Summers. 1990. "The Noise Trader Approach to Finance." *Journal of Economic Perspectives*, Spring, pp. 19–33.

Shiller, Robert J. 1981. "Do Stock Prices Move Too Much to Be Justified by Subsequent Changes in Dividends?" *The American Economic Review*, June, pp. 421–436.

Tirole, Jean. 1982. "On the Possibility of Speculation Under Rational Expectations." *Econometrica*, September, pp. 1163–1181.

Weisenbarger Investment Companies Service. 1988. *Investment Companies*, 48th Annual Edition. New York, NY: Weisenbarger Financial Services.

The Liberalization of International Trade and Payments in Eastern Europe

Norman S. Fieleke

Vice President and Economist, Federal Reserve Bank of Boston. Thomas J. Miles provided research assistance.

Few events can match the opening of the Berlin Wall as an historic symbol. Among the many things promised by that opening was the liberalization of trade that had been closely controlled for many years by the communist governments of Eastern Europe. This promise has virtually been realized in East Germany as that nation has unified with its neighbor to the West. Progress in other East European countries (including the Soviet Union) is uneven, however, because of concern over the costs of adjusting to freer trade.

This article examines the nature, motivation, and consequences of state-directed trading as it has been practiced in the centrally planned economies of Eastern Europe. Attention is then given to the issues involved in liberalization. Some general considerations suggest that state direction of foreign commerce may prove to be a tenacious legacy in at least some of the countries under consideration.

Foreign Trade under Central Planning: The Tail of the Dog

In the typical centrally planned economy, foreign trade is the tail—not the proverbial tail that wags the dog, but a more ordinary tail without much influence on the rest of the dog. The means of production are owned almost solely by the state, and central planners decide not only what will be produced by the state enterprises, but from whom the enterprises will obtain their inputs and to whom they will sell their output—and at what prices. The planners thus must balance supplies and demands for thousands of commodities. Goals are specified in terms of output quantities and are commonly unrealistically high, and prices bear little relation to those that would be set by free markets.

In this system imports are viewed more or less as necessary evils to allow fulfillment of the plan at acceptable costs. One reason for this attitude is that central planning was adopted in the first place in order to

exercise detailed control over the domestic economy, and such control is generally considered vulnerable insofar as the economy is dependent on goods from abroad. Thus, planners are loath to rely upon foreign goods unless the resource cost of domestic substitutes is substantially greater. Similarly, exports, far from being a source of pride, are perceived as a resource drain that must be endured in order to pay for imports needed to fulfill the plan.

So that it can be subjected to detailed control, the foreign trade of the centrally planned economy is carried out chiefly by state-managed foreign trade organizations. Each reports to the Ministry of Foreign Trade and has exclusive responsibility for trade in a specified range of products. The volume, commodity composition, and geographic pattern of trade to be undertaken by each foreign trade organization is specified in plans approved by the central authorities.

Because the foreign trade organization must acquire the imports indicated for it in the plan, it is not free to bargain with foreign suppliers over the aggregate amount to be purchased, although it may encourage competition among them over the price. By contrast, in marketing its exports the foreign trade organization must meet a revenue rather than a quantity goal, and might restrict the total quantity sold below that contemplated by the plan if the result were to raise the price enough to compensate for the diminished quantity.

In the typical centrally planned economy, imports and exports are viewed more or less as necessary evils to allow fulfillment of the plan.

Unlike free marketeers, central planners need to prescribe what goods will be given up to the rest of the world and what will be obtained in return. In order to regulate closely the quantities of imports and exports, and in order to assure that exports yield a desired level of imports, planners often enter into barter-like agreements and attempt to balance their trade not only worldwide but also country by country. Such barter and bilateral balancing agreements are, of course, more common in trade among centrally planned economies than in trade between centrally planned and market economies, which rely much more heavily on free markets to allocate resources.

To enforce their controls over exports and imports, central planners rely in part on controls over the use of currency for transactions relating to foreign trade. Foreign residents holding balances of a centrally planned economy currency are allowed to use them only for specified purposes. Because such foreign-held balances may not be expended for the purchase of many commodities, these balances are cursed with what is called "commodity inconvertibility." It is even more difficult for a foreigner to convert the currency of the typical centrally planned economy into freely usable currencies; thus, its currency also suffers from "currency inconvertibility."

Residents of the centrally planned economy, too, are strictly regulated in their purchases of foreign currency. Were they allowed freely to acquire foreign-currency balances, they would use those balances in part to purchase and import foreign goods, and might well spend less on domestically produced goods than projected in the central plan. Thus, foreign-currency balances accruing to foreign trade organizations in exchange for their exports must be channeled to the foreign-exchange control authority, which then allocates those balances for approved uses.

In countries that have embraced this system, the adverse consequences are plain to see. Because of the central planners' desire to minimize imports, domestic producers in these economies encounter very little import competition. This protection from foreign competition, combined with a dearth of domestic competition, allows the typical domestic producer to concentrate on satisfying the quantity goals set for it with goods that are decidedly inferior to those available in world markets. Nor is there much incentive to innovate, or to specialize in product lines for which perceived domestic demand and quantity goals are relatively low, however great the demand in the rest of the world.

Lacking the goad of competition, the typical centrally planned economy also lacks the price structure that would be set by competitive markets reflecting the underlying preferences of consumers and the true costs of production. This is not to say that such underlying preferences and costs are perfectly reflected by prices in the ordinary market economy, but the distortion is much greater in the typical

centrally planned economy. Nowhere is the distortion more obvious than in the long queues of customers seeking meat and other goods whose supply falls far short of the demand at the controlled price.

In sum, central planning seeks to manage the flow of goods and services, so planners strive to insulate their economies from foreign developments they cannot control. Thus, foreign trade and the use of currencies for foreign trade are closely regulated; prices diverge widely from those prevailing in world markets; and domestic producers experience neither the competitive pressures nor the profit incentives that exposure to foreign markets has to offer.

The Council for Mutual Economic Assistance

Imagine the centrally planned economy writ large, embracing a number of such countries, and you have something like the CMEA (Council for Mutual Economic Assistance, also known as COMECON). The CMEA was founded in 1949 by the Soviet Union, Bulgaria, Czechoslovakia, Hungary, Poland, and Romania; the German Democratic Republic joined in the following year. It is these countries on which this article focuses, although the CMEA was joined by Mongolia in 1962, Cuba in 1972, and Vietnam in 1978.

The CMEA functioned into early 1991, and oversaw trade among its members. The organization's aims, principles, functions, and powers were set forth in its charter, which is worth quoting both for its ambitious scope and for its socialist vernacular (Paxton 1989, p. 48):

> Article 1. Aims and Principles: 1 'The purpose of the Council is to promote, by uniting and co-ordinating the efforts of the member countries, the further extension and improvement of co-operation and the development of socialist economic integration, the planned development of their national economies, the acceleration of economic and technical progress in these countries, higher level of industrialization of the less industrialized countries, a continuous increase in labour productivity, a gradual approximation and equalization of economic development levels and a steady improvement in the wellbeing of the peoples.'
>
> Article 3. Functions and Powers: to (a) 'organize all-round . . . co-operation of member countries in the most rational use of natural resources and acceleration of the development of their productive forces'; (b) 'foster the improvement of the international socialist division of labour by co-ordinating national economic development plans, and the specialization and co-operation of production in member countries'; (c) 'to assist in . . . carrying out joint measures for the development of industry and agriculture . . . transport . . . principal capital investments . . . [and] trade'.

The supreme authority of the CMEA was the annual Council of prime ministers. Council decisions had to be unanimous.

Almost from its inception, the CMEA failed to pursue its professed goal of region-wide economic integration. Instead, the member countries sought a high degree of self-sufficiency, with national economic policies formulated under the supervision of the Soviet Union rather than the Council, and with trade among the members regulated by bilateral agreements. By the second half of the 1950s, it had become clear that the costs of the autarkic policies being followed were very high. Thus, CMEA members tried to breathe new life into the Council and formulate a more genuinely regional economic policy, especially by agreeing upon product lines in which each member country would specialize and by agreeing to promote the regional mobility of factors of production as well as goods. These agreements were not carried out, however, and the economic plans of the member countries were not coordinated in keeping with any formal assessment of underlying comparative advantages.

In 1971 CMEA members strengthened their avowed commitment to economic integration, agreeing to eliminate gradually the obstacles they had imposed against the free intra-regional flow of goods and services, and agreeing also to reduce barriers against the movement of productive factors. Two principal instruments were to be employed to foster integration. Foremost was to be the coordination of national economic plans while the plans were still in the draft stage. The second instrument was to be a larger role for market forces in determining prices,

Central planning seeks to manage the flow of goods and services, so planners strive to insulate their economies from foreign developments they cannot control.

interest rates, exchange rates, and the allocation of resources. Again, the rhetoric far exceeded the results, and in February 1986, Mr. Gorbachev assailed the CMEA's "armchair administration" and "endless committee deliberations." (U.N. Department of International and Economic Affairs 1989, p. 123.) Despite the criticism, by 1988 the best that the CMEA members could do was a communique reaffirming (with Romania demurring) an "earlier decision regarding the stepwise establishment of the conditions for the mutual free movement of goods, services and other production factors with the goal of creating eventually a unified market, after the preconditions thereof have been examined." (p. 128.)

What the CMEA actually embraced bore little resemblance to a free and unified market. Instead, trade among the members was closely controlled. Each country negotiated with every other member country an agreement specifying the composition and the volume of trade. Because prices within CMEA countries failed to reflect true underlying costs, the prices at which goods were exchanged were usually negotiated around a moving average of prices observed in other, freer markets. The prices so negotiated by differing pairs of countries were not identical, however.

On occasion the negotiated prices were clearly more favorable to some CMEA countries than current world prices would have been. The outstanding example is the relatively low price reportedly charged by the Soviet Union for its exports of petroleum and other raw materials to other CMEA members for many years. Soviet subsidies in this form seem to have been extended in much larger amounts to some East European countries than to others, perhaps as a reward for political allegiance to the Soviet Union. Had such rewards been paid openly as lump-sum transfers, they might have been resented by the populaces of both the Soviet Union and the East European nations (Marrese and Vanous 1988, pp. 185–220).

What the CMEA actually embraced bore little resemblance to a free market. Trade among the members was closely controlled.

It would be a grave mistake to take these subsidies as the measure of net economic benefit or cost of the CMEA to the members receiving or paying them. As already noted, centrally planned economies generally suffer from their insulation from foreign competition, and the CMEA provided such insulation. Indeed, it was a vehicle for sustaining production techniques and output decidedly inferior to those in other industrial countries (Hillman and Schnytzer n.d.).

The differences in prices—and real wages—between CMEA countries provided an incentive to shift both goods and labor from low-price or low-wage countries to the higher-price or higher-wage countries. Such shifts, which would tend to establish roughly the same price for a good throughout the CMEA, are a sine qua non for economic integration, but they would have disrupted the detailed economic plans promulgated by the authorities and, therefore, met with official resistance. Thus, controls were imposed over the very movements of goods and labor that were crucial for progress toward the professed goal of economic integration.

The shortcomings and contradictions of the CMEA were epitomized in its approach to making payments between nations and settling imbalances in international accounts. Even though central planners generally strive to avoid exporting more than is required to pay for planned imports, and thus incline to bilateral balancing, circumstances might lead a CMEA country to realize a trade surplus, even with another CMEA member. For example, one CMEA country might fail to deliver all the exports promised another during a year. CMEA procedure was for the country with the export surplus to be credited with a "transferable ruble" balance in the International Bank for Economic Cooperation (IBEC) in Moscow, while the deficit CMEA member incurred an equivalent indebtedness to the IBEC, or a reduction in its transferable ruble holdings.

The transferable ruble is considered an inferior means of payment, however. Rather than rubles, goods are what the authorities in the surplus country wanted; and the transferable ruble balance cannot be exchanged for goods without first negotiating the exchange in the form of another bilateral trade agreement, which again may go unfulfilled. Further undermining the value of the transferable ruble balance is its failure to yield interest that can be converted into merchandise any more readily than the principal.

The general failure to use a convertible currency

in CMEA transactions reinforced the tendency toward bilateral barter and blocked progress in knitting together the economies of the member countries. A CMEA country that found itself accumulating a significant transferable ruble balance had good reason to intensify its controls over trade and payments, especially its exports, in order to forestall further trade surpluses within the CMEA, while CMEA debtor countries, which were receiving essentially interest-free loans, had little incentive to increase their exports so as to repay their debts. On these counts, the CMEA tended to shrink rather than expand trade within its area.

The general failure to use a convertible currency in CMEA transactions reinforced the tendency toward bilateral barter and blocked progress in knitting together the members' economies.

Moreover, because the prices at which goods were exchanged within the CMEA were not appropriate to equate supply and demand, chronic shortages developed for some goods and chronic surpluses for others. The goods in short supply could readily be sold in world markets for convertible currencies at no discount from their CMEA prices and were dubbed "hard" goods, but goods in excess supply could be sold for convertible currencies only at a discount, often sizable, and were dubbed "soft" goods. Of the goods produced within the CMEA, fuels, food, raw materials, and various semi-manufactures typically qualified as hard goods, while other semi-manufactures and finished manufactures commonly fell into the soft goods category. Not surprisingly, CMEA members generally preferred to obtain hard goods in exchange for any hard goods that they exported, so that a tendency developed to balance not only the total trade, but trade in hard goods, between pairs of CMEA countries. This "structural bilateralism," as it was called, nicely illustrates the proclivity for one government control to beget another. The net result of such practices was that "socialist economic integration" proceeded little further than the lips of CMEA officials.

Steps toward Liberalization

The dubious value of the transferable ruble led some CMEA countries that were accumulating them to seek more nearly convertible payments for their exports to other CMEA countries and to channel more of their exports to non-CMEA countries in exchange for convertible currencies. This, however, was only the tip of the iceberg. As the 1980s drew to a close, dissatisfaction with the CMEA had become widespread and profound among the membership, as had dissatisfaction with central planning more generally. Calls were heard for radical reform or abandonment of the organization, and in January 1991 its termination was announced. Its passing should not be mourned.

For years the subject of reform had been debated within the CMEA without significant results; 1990 witnessed a turning point. Failing to agree on the recommendations tabled by a special reform commission in January 1990, the members of the CMEA appointed still another reform commission, this time to prescribe "radical" reforms. This latter commission proposed that the CMEA be downgraded into something like a regional economic secretariat, carrying out research and disseminating data, and playing little or no role in trade and payments negotiations among the members.

Also, in June 1990 CMEA officials reportedly adopted the goals of valuing their trade flows at world market prices and of settling imbalances in convertible currencies rather than in transferable rubles, although no precise timetable was promulgated. The Soviet Union is reported to have entered into agreements to start conducting its bilateral trade on this basis with the former German Democratic Republic, with Czechoslovakia, and with Hungary as early as January 1, 1991, although it is likely that the transition will require some time, with many initial exceptions to the new valuation and settlement rules.

In addition, several CMEA countries have eased, in varying degrees, their centralized control over their international trade. Producing enterprises have been allowed more latitude in selecting the foreign trade organizations with which they deal, and in some cases have even been allowed to deal directly with foreign firms. The result has been a relaxation, sometimes significant, of the state monopolization of foreign trade.

Perhaps the most publicized of these decentralizations occurred last year in the Soviet Union, where enterprises were granted widespread autonomy to

trade directly with foreign partners without intermediation of the foreign trade organizations. The international payments system employed by the country was not revamped accordingly, however, and partly because of this failing the Soviet Union soon fell deeply into arrears on its foreign debt. This incident affords but one illustration, among many that could be cited, of a much debated and very thorny problem: how to sequence the various liberalizing reforms that are required to convert a centrally planned economy into a market economy. Although the issue is too complex to be treated comprehensively in this article, we can at least explore some approaches to reform, continuing to concentrate on the realm of international trade and payments.

What Next? Some General Considerations

Many issues of economic policy are too complex to allow unambiguously correct diagnoses and prescriptions to be drawn from the corpus of economic theory and experience. And on the question of the sequencing—including the speed—of liberalizing reforms, would-be analysts and policymakers soon discover that the accumulated knowledge to which they can appeal is extremely limited. As was stated in a recent publication of the International Monetary Fund (IMF 1990, May, p. 70), "There is no theory of the transition from a centrally planned to a market economy. Nor are there yet examples of centrally planned economies that have successfully made such a transition."

With so little guidance from both theory and experience, it seems appropriate to begin with the hypothesis that liberalization should be introduced quickly rather than being phased in gradually, sector by sector. If it is true, as widely believed, that centrally planned economies have suffered from the government controls that pervade their economies, the remedy would seem to be to relax those controls immediately, just as one would remove a straitjacket. The IMF makes the case as follows (p. 71):

> For a number of reasons, a rapid implementation of market-oriented reforms may be preferable to a gradual approach. The more rapid the reforms, the less those who benefited from the old system and other interest groups will be able to obstruct or slow the implementation of the reform program. In addition, public consensus in support of reform . . . can best be maintained by ensuring that the tangible benefits of reform become visible as soon as possible. There is no convincing argument that the transitional costs of structural reform would be reduced if the reform process were prolonged or delayed . . . .
>
> The transitional costs . . . will . . . depend on the expectations of market participants . . . and on the credibility of the policy itself . . . . The credibility of an economic reform package, and hence the probability of it succeeding, is likely to be greater if it is comprehensive. Given the linkages in the economy, comprehensive reform increases the likelihood that each element of the program will reinforce other elements. Moreover, comprehensive reform helps ensure that the costs and benefits of economic transformation are broadly shared rather than concentrated on specific segments of society . . . .
>
> In practice, of course, everything cannot be done at once, and even a rapid approach to reform will involve short-run choices concerning the pace of implementation of specific measures . . . . It would seem to be particularly important . . . that the establishment of macroeconomic stability and institutional changes such as modifications to the legal system, the creation of social safety nets, and the establishment of financial discipline on [government] enterprises be emphasized at an early stage.

This is one of the most cogent and succinct statements of the case for rapid liberalization. From this and other arguments, it seems clear that strong public support will be a crucial condition for success. Such support must extend to tolerance of occasional errors committed by well-intentioned policymakers who have the courage to embark on these uncharted waters, and tolerance of at least temporary reductions in income for many who will become exposed to more intense competition as government intervention is reduced.

To forestall harsh opposition from those disadvantaged by the reforms, the beneficiaries may have to contribute heavily to unemployment benefits, retraining programs, and the like. Even in wealthy market economies, it is uncommon for significant income-redistributing liberalizations to be introduced precipitously and without compensation for those distressed. Tariff reductions, for example, are generally phased in over several years so as to allow the adversely affected to make an orderly adjustment, and—at least in the United States—have been accompanied by generous adjustment assistance.

While most liberalizing countries may well decide against the immediate adoption of free trade, an early and substantial relaxation of controls over international trade and payments is probably critical to the rapid development of competitive markets. It is likely

to take some time for markets to flourish within these countries and to establish relative prices that channel resources with great efficiency. In particular, monopolistic government enterprises cannot be broken down into smaller units and converted to private ownership in the twinkling of an eye. In the meantime enhanced foreign competition can exert a salutary discipline on these enterprises.

The opening up of trade will introduce the competitive price structure—or set of relative prices—that predominates for goods traded throughout the rest of the world. The importance of this reform can hardly be exaggerated. A key impediment to economic progress within the CMEA was an inappropriate allocation of resources (including, of course, human effort), as central planners failed to perform this allocative function as well as competitive markets. It is world prices to which the liberalizing economies must adapt if they are to reap the benefits of integration with the international economy.

The Issue of Convertibility: A Payments Union?

As already noted, it has been the practice within centrally planned economies to restrict the use of the domestic currency, both for the purchase of domestic goods and for the purchase of foreign currencies. Until these countries allow their currencies to be exchanged freely for goods and for foreign currencies, their domestic prices will fail to reflect world prices, and the inefficiencies associated with this failing will persist. In addition to the inefficiencies already mentioned, foreign investment in the centrally planned economies has been greatly discouraged by the obstacles that prevent investors from earning and repatriating profits in convertible currencies.

In the typical country whose currency is inconvertible, many domestic prices are lower than would be the case with convertibility and efficiency. Indeed, one of the purposes of inconvertibility is to assist the government in suppressing prices below their free market levels. By restricting the use of the domestic currency to purchase foreign exchange, the authorities limit any bidding up of the price of foreign exchange, and of the domestic currency price of foreign goods, by those who would like to enter the market. This practice, of course, contributes to the widespread shortages, queues, and black markets.

Suddenly to allow convertibility and the associated free-market determination of prices would be to risk a sharp depreciation of the country's currency in the foreign-exchange market (or rapid exhaustion of the foreign-exchange reserves that the government could supply to the market in an effort to prevent the depreciation), a quantum leap in the general price level, and a marked rise in unemployment as the changing prices rendered many lines of activity less viable. These consequences are not idle fears but have in fact materialized in more than one country undertaking liberalization.

Seeking ways to minimize such transitional costs, many analysts have stressed the importance of pursuing anti-inflationary policies and of reforming the laws on private property, taxation, and commerce to provide a seedbed for private enterprise prior to any substantial liberalization. Others, doubting the adequacy of such measures, have proposed that the East European countries should ease the transition to freer international trade and payments by following the example of West European countries after World War II: they should form a payments union designed not only to conserve their scarce foreign-exchange reserves but also to transform their system of predominantly bilateral balancing into one of multilateral balancing, with the transformation occurring gradually over a period of years and culminating in full currency convertibility. The balance of this paper examines this proposal and finds it wanting, in spite of its cosmetic appeal.

Eastern and Western Europe: Historical Analogies

The reference to postwar Western Europe is intriguing, because the situation of the East European countries today is in many respects similar to that of their West European neighbors shortly after World War II. At controlled prices and exchange rates, shortages were widespread in Western Europe. To insure that their limited supplies of convertible foreign exchange would be used to acquire goods deemed essential, governments in the area generally exercised tight control over international trade and payments. Trade carried on by government monopolies could, of course, be regulated directly, just as in the centrally planned economies. Trade by private parties was controlled through the issuance of licenses authorizing the exportation or importation of specified quantities or values of merchandise.

International payments were regulated by ex-

change controls administered through the banking system. Businessmen were to sell to their domestic banks the foreign exchange they earned from foreigners, and they might buy foreign exchange from the banks for authorized purposes. If a commercial bank ran low on foreign exchange demanded by customers for approved transactions, the bank could acquire more at the country's central bank, while a commercial bank that accumulated excess foreign exchange could sell it to the central bank.

In order to avoid incurring deficits on their international transactions that would have drained their foreign-exchange reserves and stifled their trade, the countries of Western Europe entered into bilateral trade and payments agreements with one another, just as the CMEA countries did. Such an agreement specified the trade to be permitted between the two signatories and the exchange rate between their currencies.

Of course, in the event, trade between the parties to these agreements was less than perfectly balanced, and as a practical matter could not have been balanced on a daily basis. To allow for such imbalances, each central bank maintained an account with overdraft privileges at the other country's central bank. A central bank that was exhausting its stock of the other country's money could draw an overdraft on the other's central bank; the drawing bank would then credit an equivalent amount in its own currency to the foreign central bank. These overdrafts provided a "swing" that accommodated temporary imbalances, just as additional governmentally owned foreign exchange would have done.

The situation of the East European countries today is in many respects similar to that of their West European neighbors shortly after World War II.

A central bank on which overdrafts were drawn was, of course, extending credit, accepting in return deposit balances in the other country's currency. As these balances mounted, negotiations would commence on how the imbalance in trade might be eliminated or reversed or on how much of the debtor country's indebtedness (that is, of its currency held by the creditor's central bank) would be paid off in gold or in some currency acceptable to the creditor. To avoid accumulating balances of inconvertible currencies, a West European country was inclined to discriminate in favor of imports from the countries with which it was running trade surpluses. Again, the parallel with the CMEA is clear.

The European Payments Union

In an effort to reduce this incentive for bilateral balancing, the West European nations embarked on an historic international financial arrangement known as the European Payments Union (EPU). Established in September 1950, the EPU functioned through December 1958. Because it fostered the revival of multilateralism in trade and payments among the countries of Western Europe, it became a model for proposed regional payments arrangements among countries with inconvertible currencies—including, most recently, an arrangement for Eastern Europe.

The rules and procedures of the EPU were complex and were modified as time went by, and only some of the salient features are summarized here. The cornerstone of the EPU was multilateral rather than bilateral settlement of payments imbalances among the members. As before, each member's central bank stood ready to lend its own currency to other member central banks in order to satisfy the demand for that currency at the agreed exchange rates. Each month the net amount of such lending or borrowing by every member vis-à-vis all the other members as a group was tallied by the Bank for International Settlements and recorded as a claim on or debt to the Union, expressed in an agreed unit of account. This procedure economized on the use of scarce foreign-exchange reserves and diminished the proclivity for bilateral balancing, since each country could offset a deficit with another country or countries with any surpluses it might have with still other countries. Confidence was inspired by the fact that each country's net claims or net debt were now with the Union rather than with other individual countries, meaning that credit risks were assumed by the group as a whole. Interest was paid to creditors and collected from debtors at rates rising with the duration of the debt.

Once a country's net debt to or claims on the

Union exceeded a limit calculated by an agreed formula, the country was to make or receive settlement in gold or dollars for at least part of the excess. As time passed the requirement for settlement in gold or dollars was stiffened, a modification that put greater pressure on chronic deficit countries to reduce their overall deficits and also reduced the incentive to discriminate against imports from non-Union countries settling payments in convertible currencies, primarily the United States.

It is noteworthy that the liberalizing intent of the West European countries was manifested in 1950 not only by the establishment of the EPU but also by the initiation of a program to reduce nonmonetary barriers to trade within Western Europe. In October of that year the West European countries agreed to eliminate quantitative restrictions in a nondiscriminatory fashion from at least three-fifths of their imports from one another. Import quotas were further relaxed in subsequent years.

Despite the good intentions, the transition to convertibility was neither rapid nor uninterrupted. Early on, both the trade liberalization program and the credit facilities of the EPU were tested by large payments imbalances. The West European countries met the challenge largely by increasing the amounts to be loaned and borrowed under the aegis of the EPU, and by adapting the pace of trade liberalization (reversing it for deficit countries and accelerating it for surplus countries), thereby buying time for other balance-of-payments adjustment measures to take effect. Also, the claims on and debts to the Union of some countries were converted into claims on and debts to other individual members. Without such flexibility the EPU would surely have foundered.

Progress toward convertibility was facilitated not only by this flexibility but also by U.S. aid and large overall deficits in the U.S. balance of payments. The counterpart of much of these deficits was an increase in the gold and dollar reserves of West European central banks. As these reserve stocks mounted, their holders became more disposed to making sales from them in exchange for their native currencies.

Thus, some eight years after the founding of the EPU—on December 27, 1958—the major step toward formal convertibility was at last taken. Belgium-Luxembourg, France, Italy, the Netherlands, the United Kingdom, and West Germany, soon followed by other West European nations, announced that their currencies would be convertible for foreign residents. A nonresident of any one of these countries who earned its currency in a current-account transaction (such as exporting) could thereafter freely sell that currency in exchange for any other currency, including dollars, at the officially supported rates of exchange. Having performed its function, the EPU was then terminated.

A Model for Eastern Europe?

From this capsule review, it is clear that the EPU was successful, albeit slow, in dealing with some of the same maladies that have afflicted the East European countries, although the dislocations within the EPU economies originated chiefly in a devastating war rather than in the failures of central planning. Maladies common to both sets of countries have included a perceived shortage of convertible foreign-exchange reserves, bilateral balancing of international trade, foreign-exchange rates and internal prices distorted by controls, and inconvertible domestic currencies. If, as widely believed, the EPU materially assisted its members in overcoming these ailments, should the East European countries establish a similar union?

Now the erstwhile CMEA members are liberalizing at widely different paces; to combine them into a payments union would be to run a high risk of slowing the pace of overall liberalization.

Despite the similarities between some of the problems faced by the countries of the EPU and Eastern Europe, closer scrutiny reveals differences that raise serious doubts about the usefulness of a payments union for Eastern Europe. To begin with, the EPU embraced a much wider trading area, and accounted for a much greater share of the members' total trade, than is true of the East European countries likely to favor a payments union. Proposals for an East European Payments Union (EEPU), have generally contemplated including Bulgaria, Czechoslovakia, the German Democratic Republic, Hungary, Poland, Romania, and the Soviet Union. Now, of

course, the German Democratic Republic is no more; and the Soviet Union might well remain outside an EEPU, largely because, as a member, it would likely encounter requests for substantial ongoing credit from other members, especially as they begin to pay for Soviet oil in convertible currency at something like world prices.

Moreover, as the record clearly shows, past actions and relationships within the CMEA would be a poor foundation on which to begin the construction of a liberal trade and payments regime. In particular, a multilateral payments system could have been developed within the CMEA around modifications of the International Bank for Economic Cooperation (IBEC) and the transferable ruble, but the opportunity went unexploited. The IBEC could have played a role similar to that of the Bank for International Settlements in facilitating multilateral rather than bilateral settlements, and the transferable ruble could have been made truly transferable among CMEA members in exchange for goods. Now the erstwhile CMEA members are liberalizing at widely differing paces, and some have already introduced a high degree of currency convertibility. If history offers any guide, to combine them into a payments union would be to run a high risk of slowing the pace of overall liberalization to that of the most reluctant members.

Crucial to a rapid transition to a relatively free market economy is the adoption of liberal foreign trade and payments arrangements, including a high degree of currency convertibility.

Finally, during the years of the EPU, central banks generally strove to restrain exchange-rate movements within very narrow ranges, while today many exchange rates are allowed to move much more freely under the influence of market forces. Were the East European countries to tolerate relatively free movement of their currency exchange rates in open markets, they would both establish convertibility for their currencies and obviate the need for complex and dubious transitional mechanisms such as payment unions. Of course, for exchange rates to settle at levels that reflect domestic market equilibrium prices, prices within Eastern Europe must be substantially freed from controls—as some of the countries have recently done—and measures supporting the development of free markets must be promptly introduced. To ease this difficult transition, industries that are particularly vulnerable to newly encountered foreign competition might be granted temporary tariff protection, to be phased out according to a well-publicized schedule.

Conclusion

Current experience is demonstrating that the transition from a centrally planned to a relatively free market economy is far from costless. However, the cost represents an investment that should yield immense returns in the longer run. Crucial to a rapid transition is the adoption of relatively liberal foreign trade and payments arrangements, including a high degree of currency convertibility.

A clear and present danger is that the countries undertaking the transition will fail to follow through as the adjustment costs materialize. State enterprises and controls have provided a livelihood for many who will resist their demise. Nonetheless, market-oriented systems will probably be adopted eventually in view of their demonstrated superiority, and countries that falter in their reforms may therefore merely prolong their agony.

References

Bofinger, Peter. 1990. "A Multilateral Payments Union for Eastern Europe?" London: Centre for Economic Policy Research, August, Discussion Paper No. 458.

Central Banking Issues in Emerging Market-Oriented Economies. Kansas City: Federal Reserve Bank of Kansas City.

Deutsche Bank, Economics Department. 1990. *Special Eastern Europe*, Otto Sorf, ed., February 7. Frankfurt: Deutsche Bank.

Diebold, Jr., William. 1952. *Trade and Payments in Western Europe: A Study in Economic Cooperation*. New York: Harper & Brothers for the Council on Foreign Relations.

Fieleke, Norman S. 1990. "Commerce with the Newly Liberalizing Countries: Promised Land, Quicksand, or What?" *New England Economic Review*, May/June, pp. 19–33.

Friedman, Milton. 1990. "Four Steps to Freedom." *National Review*, May 14, pp. 33–36.

Hillman, Arye L. and Adi Schnytzer. Undated. "Creating the Reform-Resistant Dependent Economy: An Analysis of the CMEA International Trading Relationship." Bar-Ilan University, Israel, unpublished paper.

Holzman, Franklin D. 1985. "A Comparative View of Foreign Trade Behavior: Market Versus Centrally Planned Economies." In Franklin D. Holzman, *The Economics of Soviet Bloc Trade and Finance*. London: Westview Press, 1987.

Hudson Institute, Inc. 1990. *Hungary in Transformation to Freedom and Prosperity*. Economic Program Proposals of the Joint Hungarian–International Blue Ribbon Commission, April. Indianapolis: Hudson Institute, Inc.

Institute of International Finance, Inc. 1990. *Building Free Market Economies in Central and Eastern Europe: Challenges and Realities*, April. Washington, D.C.: Institute of International Finance, Inc.

International Monetary Fund. 1990. *World Economic Outlook*, May and October 1990. World Economic and Financial Surveys. Washington, D.C.: IMF

International Monetary Fund *et al*. 1990. *The Economy of the USSR: Summary and Recommendations*. Washington, D.C.: IMF.

Kaplan, Jacob J. and Gunther Schleiminger. 1989. *The European Payments Union: Financial Diplomacy in the 1950s*. New York: Clarendon Press.

Kenen, Peter B. 1990. "Transitional Arrangements for Trade and Payments Among the CMEA Countries." Unpublished International Monetary Fund Working Paper, European Department, September.

League of Nations, Economic, Financial, and Transit Department. 1943. *Trade Relations between Free-Market and Controlled Economies*. Geneva: League of Nations.

Lipton, David and Jeffrey Sachs. 1990. "Creating a Market Economy in Eastern Europe: The Case of Poland." In *Brookings Papers on Economic Activity 1 1990*, William C. Brainard and George L. Perry, eds., pp. 75–147. Washington, D.C.: The Brookings Institution.

Marrese, Michael. 1986. "CMEA: Effective but Cumbersome Political Economy." *International Organization*, vol. 40, no. 2, pp. 287–327.

Marrese, Michael and Jan Vanous. 1988. "The Content and Controversy of Soviet Trade Relations with Eastern Europe, 1970–1984." In *Economic Adjustment and Reform in Eastern Europe and the Soviet Union*, Josef C. Brada, Ed. A. Hewett, and Thomas A. Wolf, eds., pp. 185–220. Durham: Duke University Press.

Paxton, John, ed. 1989. *The Statesman's Year-Book, 1989–90*. New York: St. Martin's Press.

Schrenk, Martin. 1990. *The CMEA System of Trade and Payments: Today and Tomorrow*. Strategic Planning and Review Discussion Paper No. 5, January. Washington, D.C.: The World Bank.

Triffin, Robert. 1957. *Europe and the Money Muddle: From Bilateralism to Near-Convertibility, 1947–1956*. New Haven: Yale University Press.

United Nations, Department of International Economic and Social Affairs. 1989. *World Economic Survey 1989. Current Trends and Policies in the World Economy*. New York: United Nations.

———. 1990. *World Economic Survey 1990: Current Trends and Policies in the World Economy*. New York: United Nations.

United Nations, Secretariat for the Economic Commission for Europe. 1989. *Economic Bulletin for Europe*. New York: United Nations.

———. 1990. *Economic Survey of Europe in 1989–1990*. New York: United Nations.

Van Brabant, Jozef M. 1987. *Adjustment, Structural Change, and Economic Efficiency: Aspects of Monetary Cooperation in Eastern Europe*. New York: Cambridge University Press.

———. 1990. "On Reforming the Trade and Payments Regimes in the CMEA." Draft paper, May 31.

Wolf, Thomas A. 1988. *Foreign Trade in the Centrally Planned Economy*. London: Harwood Academic Publishers.

———. 1990. "Market-Oriented Reform of Foreign Trade in Planned Economies." Unpublished International Monetary Fund Working Paper, European Department, April.

Yeager, Leland B. 1976. *International Monetary Relations: Theory, History, and Policy*. New York: Harper & Row.

Financing Capital Expenditures in Massachusetts

Alicia H. Munnell and Leah M. Cook

Senior Vice President and Director of Research, and Research Assistant, respectively, Federal Reserve Bank of Boston. This article first appeared as a chapter in Federal Reserve Bank of Boston Research Report No. 72, Massachusetts in the 1990s: The Role of State Government. *The authors' acknowledgments appear at the end of this article.*

Spending on capital projects in Massachusetts has not contributed in any significant fashion to the state's budget crisis. Rather, this state, like others, has probably spent too little, and not too much, on public infrastructure. The reasons for underspending are clear. The states nationwide are caught between the increased requirements of localities and decreased funding from the federal government. The result everywhere is infrastructure in disrepair and a host of unmet capital needs.

The Massachusetts situation is particularly troublesome. The state spent most of the 1980s embroiled in conflict with the Administration over federal funding for the Central Artery Depression/Third Harbor Tunnel project. During the same period, the state's federal grants for other projects dropped sharply and the condition of much of the Commonwealth's infrastructure deteriorated markedly compared to the rest of the nation. This deterioration occurred during a period of rapid economic growth, which placed enormous demands on public capital, and of rising tax revenues, which could have been used for public capital investment. Now the condition of Massachusetts' public capital can be best described as average. Belatedly, federal funding for most of the Central Artery/Tunnel project has been appropriated and activity is underway. But these initiatives coincide with a serious downturn in the state's economy, which raises questions about the state's ability to come up with its share of the financing. More generally, if funding for infrastructure other than the Central Artery/Harbor Tunnel project did not materialize during the boom, where will the money come from during the 1990s? These are the issues addressed in this article.

The first section briefly sets out the role for the state in financing physical infrastructure and compares that role to the present division of responsibilities among the federal, state, and local governments. It also describes the major federal grant programs for capital investment.

The second section assesses Massachusetts' ability to meet its

infrastructure needs. It compares capital spending here with that in comparable states. It also examines the effectiveness of that spending by looking at the physical condition of roads and bridges in Massachusetts and other states. New data on state-by-state public and private capital provide an opportunity to compare the net effect of investment in Massachusetts with that undertaken elsewhere.

The third section examines the budgeting and decision-making process that produces the current level of capital expenditure, and explores both the capital budgeting process of the state government and the role of public authorities in capital spending.

The fourth section summarizes the state's capital spending plans for the 1990s. This includes a discussion of how the "mega" projects—the Central Artery Depression/Third Harbor Tunnel project and the Boston Harbor Cleanup—fit into the Commonwealth's infrastructure agenda. It also highlights the pressure that these projects and the unmet capital needs of the 1980s will put on the budgets of the 1990s.

The article concludes that Massachusetts' current complex and ambitious capital spending agenda requires centralized decision-making and a mechanism for ranking projects by their importance. The current fragmentation of initiatives and financing among the state and independent authorities is no longer workable.

I. The Role of the State in Financing Public Infrastructure

The basic rationale for government financing of capital expenditures is that some necessary and desirable investments would not be undertaken if left to the private sector. The problem arises because certain capital projects immediately provide benefits to everyone in a town or state or nation, as soon as they are available to one person. A typical example is a park. The benefits of a park cannot be divided up and meted out only to those willing to pay. The inability to exclude those unwilling to pay means that a profit-seeking builder would have no incentive to construct such a project.

Sometimes government provision is called for even if exclusion is possible. For example, it might be possible for a private entrepreneur to build a park with a fence around it and admit only those persons who paid an entrance fee. Such a setup would be extremely inefficient, however, since parks, as well as bridges or roads, produce services with enormous economies of scale. The initial fixed cost might be quite large, but the marginal cost of providing one more entry, crossing, or road trip is nearly zero. Therefore, excluding those unwilling to pay would simply deny some individuals the enjoyment of a service that costs next to nothing to produce.

The level of government that should undertake a particular capital investment depends on the nature of the project and the location of the people who will benefit. A small park, local road, hospital, or police station, which will be used primarily by residents of a town, should be financed by the individual locality. If the benefits of a capital investment spill over to several communities, then either the towns will have to band together or a higher level of government will have to participate in the financing in order to ensure that adequate money will be devoted to the project. A clear example is an urban beltway, which benefits residents of all the surrounding towns. If benefits spill over from one state to another, as with the interstate highway system, then a role exists for financial contributions from the federal government.

Broadly speaking, the share of total costs paid by the higher level of government should correspond to the share of total benefits enjoyed by those who live outside the participating jurisdictions. Thus, the construction of a local police station should be paid for by the residents of the local community served by the station. The urban beltway could be financed by a regional special district composed of the towns that encircle the central city; if other residents of the state benefit from the beltway, the state could provide some matching support to the district. Segments of the interstate highway system should be financed jointly by the states and the federal government, with the federal share reflecting the use of the highway by out-of-state drivers.

The level of government that should undertake a particular capital investment depends on the nature of the project and the location of the people who will benefit.

Efficient allocation also requires that, whenever possible, the construction of public capital be financed by bonds with approximately the same maturity as the expected life of the project, and that debt service of public capital investments be financed by fees or taxes that come as close as possible to user charges. Thus, if the police station has a 20-year life, it should be financed with bonds that match the real costs to the taxpayers with the stream of benefits. Explicit user charges would not be sensible for a police station, but the servicing of the bonds should be paid out of a local revenue source, such as the property tax, so that the burden falls on those individuals who benefit from the service. The construction of the beltway, again, would be financed by debt, and the interest and principal payments should be derived from a source that reflects use of the road—registration fees for automobiles in the surrounding towns might be one source. Servicing debt for the construction of the interstate highway could come primarily from gasoline taxes.

In the real world, who pays for public infrastructure, and where does the money come from? States and localities undertake almost all spending on nonmilitary public capital investment. In 1989, they spent $112 billion on highways, buildings, water supply and treatment facilities, utilities, and other initiatives (Table 1).[1] Of this total, the federal government supplied $24 billion, or 22 percent.[2]

Table 1
Public Capital Expenditures by All State and Local Governments, Fiscal Year 1989

Function	Capital Expenditures (Billions)	Percent Financed by Federal Grants
Total	$111.8	22
Transportation		
Highways	33.9	39
Mass Transit	4.3	62
Aviation	3.1	37
Water	1.0	0
Schools and Hospitals	23.6	0
Wastewater Treatment	8.9	27
Water Supply	6.5	1
Other Utilities	5.0	0
Other	25.5	19

Note: Total capital outlays were available for 1989 but outlays by function had not yet been tabulated by the Census Bureau. Capital outlays by function were estimated for 1989 by calculating the share of total expenditures by function in 1988, and applying this share to the 1989 total capital outlay.

Source: Congressional Budget Office Infrastructure Database; U.S. Bureau of the Census, 1989, *Government Finances in 1987–88*, Table 24; U.S. Bureau of the Census, 1990, *Government Finances: 1988–89 (Preliminary Report)*; U.S. Office of Management and Budget, 1990, *Budget of the United States Government, Fiscal Year 1991, Historical Tables*, Tables 9.5 and 12.3.

The Federal Role

The federal government's contribution toward financing nonmilitary capital expenditures is directed primarily at achieving transportation and, more recently, environmental objectives. The government's commitment to transportation stems from its responsibilities to promote interstate commerce and provide for national defense. Environmental responsibilities arise because the harmful effects of pollutants spread beyond the localities or states that house the polluters, and therefore would not be corrected without intervention from a higher level of government.

Most of the federal money is distributed by the Federal Highway Administration, which runs six major highway programs for construction and, to some extent, rehabilitation. While most of the money goes to the federal-aid highway system (through five separate programs aimed at different levels of roads), the Highway Bridge Replacement and Rehabilitation Program also provides some funds for the nonfederal highway system. For projects that qualify under any of these six grant programs, the federal government contributes between 75 and 90 percent of the total costs. (See Appendix A for the specific provisions of the major grant programs.)

These grants are paid from the Highway Trust Fund, created by Congress in 1956 to finance the ambitious plan to build the interstate and national defense highway system. It is funded by numerous excise taxes derived from transportation activity. Most of the money comes from a fuel tax of 9 cents (14 cents beginning December 1990) per gallon for gasoline, 3 cents (8.6 cents beginning December 1990) for gasohol, and 15 cents (20 cents beginning December 1990) for diesel fuel, but sales taxes on tires and trucks, as well as truck usage taxes, also support the trust fund. One cent of the gasoline tax is dedicated to the Urban Mass Transit System and maintained as a separate account within the Highway Trust Fund.

Funding for aviation capital expenditures is channeled through the Airport and Airways Trust Fund, established in 1970 and financed by excise taxes on passenger ticket sales, freight charges, and aircraft fuel. The matching percentages for airport capital construction range from 75 percent for the

largest airports to 90 percent for the rest, and vary by type of project.

The major grants for wastewater treatment began with the Clean Water Act of 1972, when the federal government first assumed responsibility for controlling water pollution. The Act required the Environmental Protection Agency (EPA) to establish minimum standards for municipal and industrial wastewater treatment and significantly increased federal funding. The grants are distributed to the states based on population and EPA standards of need; the states then allocate funds to local communities for building or improving publicly owned treatment facilities.[3] The program was never intended to be permanent and, as the result of amendments enacted in 1987, municipal construction grants will be phased out by 1991. Until 1994, a temporary federal program will provide seed money on a matching basis ($.20 state money for every $1 of federal funds) to establish state revolving funds, which will provide loans for future construction.

All the grant programs require periodic reauthorization, when Congress establishes a dollar cap on the amount of funds that can be appropriated for matching grants.[4] Thus, the current system has caps to limit use, but also very high matching rates that stimulate use. The matching rates are probably much higher than can be justified by any spillover effect. The Department of Transportation estimates that only 30 percent of the drivers on the average interstate highway are out of state (Gramlich 1990). The large subsidy provides an enormous windfall to states with high levels of capital spending on highways, while the cap means that at the margin the federal government provides no incentive for states to spend on federal roads. Economists have suggested that a more efficient solution would be to return funds to the states in the form of uncapped matching grants with a matching rate of 30 percent, rather than the current system of capped 90 percent grants (Gramlich 1990). The Administration is moving in this direction with its proposals for the 1992 reauthorization of the Surface Transportation Act, which significantly reduce the matching rates for highway and mass transit grants.

The goal of this review, however, is not to reform the federal system, but rather to highlight the decline in the federal government's contribution to public capital investment. This is shown clearly in Table 2, which reports expenditures (in 1989 dollars) under the major capital grant programs since 1970. The decline in the level of subsidy, which began in 1980, is projected to extend into the 1990s under the most recent federal budget proposals. Most of the decline to date reflects the phasing out of the wastewater treatment grant program, but the bulk of the projected drop rests on the sharp reduction in matching grants for highway construction under the Bush Administration's new transportation plan (U.S. Department of Transportation 1990). Mass transit grants are also projected to decline.

Moreover, judging the performance of any effort on the basis of absolute dollar amounts, even if adjusted for inflation, does not provide an accurate picture of trends in a growing economy. If the grants are measured as a percent of total state and local capital expenditures or of total federal budget outlays, the decrease in federal government support becomes even more striking. Between 1980 and 1989, federal capital grants declined from 36 percent to 22 percent of total state and local capital spending. Grants as a percent of total federal outlays fell from

Table 2
Federal Capital Grants to State and Local Governments by Function, Selected Fiscal Years
Billions of 1989 Dollars

Function	1970	1980	1989	1995 (est.)
Total	23.4	29.6	24.4	16.9
Transportation				
Highways	14.3	11.8	13.2	9.7
Mass Transit	0.5	2.7	2.7	1.4
Aviation	0.3	0.8	1.1	1.2
Wastewater Treatment	0.5	5.9	2.4	1.0[a]
Water Supply	0.3	0.3	0.1	0.1
Other	7.4	8.1	4.9	3.5
Addendum:				
Federal Grants as a Percent of State and Local Capital Expenditures	23.7	35.7	21.8	n.a.
Total Federal Outlays	3.6	3.4	2.1	1.4

n.a. = not available.

[a]Current legislation requires that the wastewater treatment grant program administered by the Environmental Protection Agency be phased out by 1991. Spending continues beyond this point, however, because these construction grants represent multiyear commitments, which are paid out over the construction period. Thus, grants to be paid out in 1995 represent spending that was authorized prior to the phaseout of the program.

Source: Congressional Budget Office Infrastructure Database; U.S. Office of Management and Budget, *Budget of the United States Government, Fiscal Year 1991* and *Fiscal Year 1989, Historical Tables*, Tables 9.5 and 12.3.

3.4 to 2.1. In short, the federal government has been quietly shifting responsibility for capital spending from the federal government to the states and localities for nearly a decade. With the Bush Administration's new transportation policy the shift now has become an explicit goal of the federal government.

The Local Role

Local governments have been the traditional providers of public infrastructure; they have built the schools, hospitals, police stations, sidewalks, and local streets. These governments historically have relied on the local property tax to support not only capital projects, but also the vast array of services supplied by cities and towns. The property tax provided an ample supply of money for local government initiatives and served as a nice proxy for both ability to pay and benefits received.

Local governments can no longer rely on obtaining adequate revenue from the property tax. Repeated property tax increases to support public services and economic development for a growing population have met with serious taxpayer resistance. This resistance has frequently culminated in state initiatives that place limits on local taxes. Even before 1970, local jurisdictions in 25 states faced limits on the tax rates they could impose on local property owners; eight more states had set limits by 1985 (ACIR 1987). California's Proposition 13 and Massachusetts' own Proposition 2½ are the best known.

Federal grants to localities have declined at the same time that state limits have been imposed on local property taxes.

At the same time that revenues have been limited, demands on the fiscal capacities of cities and towns have increased. The costs of education, law enforcement, and low-income housing have all continued to soar. Simultaneously the federal government, through the EPA, promulgated new environmental standards that significantly increased both capital and operating expenses for localities. Lack of funds led many cities to postpone both rehabilitation of old plants and new construction, only to find that the costs of these legally mandated improvements have skyrocketed. New York, Boston, and other large cities face huge infrastructure maintenance deficits and major costs to upgrade outdated wastewater treatment facilities to meet EPA standards.

Localities also suffered from the cutback in a number of federal government programs for which they were the major beneficiaries. These include revenue sharing, grants provided for low-income housing, and funds for the construction of wastewater treatment facilities.

In short, federal government grants to localities have declined at the same time that state limits have been imposed on local property taxes. Yet responsibilities of the cities and towns for schools, hospitals, and police, as well as new federally mandated environmental projects, have all increased. Given these pressures, the cities and towns have turned to the only possible source of support—the states.

II. Massachusetts' Capital Spending

How has Massachusetts performed in the changing environment? How much money has Massachusetts been spending on public infrastructure? And what is the state of repair of its capital investments?

Before looking at the data, consider the importance of public capital investment. Public investment has received much attention recently as stories abound of deteriorating public capital, especially roads and bridges. Two of the largest public capital investment projects in the state's history—the Central Artery Depression/Third Harbor Tunnel and the Boston Harbor Cleanup—have also sparked interest in the potential benefits that new construction activity might bring to a declining economy. Once built, public capital investment continues to be important to the economic vitality of the state, because it affects the locational decisions of households and firms and the productivity of businesses.

Although most observers acknowledge that public infrastructure has a positive impact on economic activity, only in the past few years have economists included measures of public capital explicitly in their models of productivity and growth. Work by Aschauer (1989) showed a strong relationship between output per unit of private capital and the stock of public capital. Munnell (1990a), examining the labor productivity slowdown in the 1970s, found a similarly strong relationship between the nation's stock of

public capital and the level of productivity growth.

These nationwide results were confirmed at the state level in a recent study that examined the impact of public capital on output, employment growth, and private investment on a state-by-state basis (Munnell 1990b). The results clearly showed that those states that have invested more in infrastructure tend to have greater output, more private investment, and higher employment growth.

Given the economic importance of public capital, how does Massachusetts measure up? The answer is that, no matter how spending or quality is measured, Massachusetts today appears more or less average. Table 3 shows the most recent data for Massachusetts capital expenditure by function; in terms of the relative shares in each category, Massachusetts looks very much like the nation as a whole (Table 1). The table also compares Massachusetts and the nation in terms of the percentage of each expenditure that is financed by the federal government. Again, the figures are very close; the only noticeable discrepancy is in the area of mass transit, where Massachusetts appears to receive a much lower rate of federal contribution than the rest of the nation.

Table 3
Public Capital Expenditures by Massachusetts' State and Local Governments and Federal Government Contributions, Fiscal Year 1989

Function	Capital Expenditures (Millions)	Percent Financed by Federal Grants	
		Massachusetts	U.S. Total
Total	$2,803.7	20	22
Transportation			
Highways	613.0	38	39
Mass Transit	426.7	25	62
Aviation[a]	a	a	37
Schools and Hospitals	357.7	0	0
Wastewater Treatment	339.5	27	27
Water Supply	116.7	1	1
Other Utilities	45.8	0	0
Other[a]	904.4	13[b]	19

Note: Total capital outlays were available for 1989 but outlays by function had not yet been tabulated by the Census Bureau. Capital outlays by function were estimated for 1989 by calculating the share of total expenditures by function in 1988, and applying this share to the 1989 capital outlay.

[a]The Census Bureau lists capital outlays on aviation for all state and local governments, but does not provide the same information for individual states. Hence, Massachusetts' expenditures for aviation are included in the "Other" category.

[b]The percent financed by federal grants could be somewhat higher because for a large component of the "Other" category it is not possible to identify the grants received by Massachusetts. The reason for this problem is that the source for grants by state is organized by agency and program, while the source for total grants for physical capital investment is organized by broad functional area. In most cases it is possible to match these two sources. Much of the "Other" category of total grants, however, includes no listing by program, which makes it impossible to estimate the grants in this category received by individual states.

Source: U.S. Bureau of the Census, 1989, *Government Finances in 1987–88*, Table 24; U.S. Bureau of the Census, 1990, *Government Finances: 1988–89 (Preliminary Report)*; U.S. Bureau of the Census, 1990, *Federal Expenditures by State for Fiscal Year 1989*, Table 2.

Table 4
Federal Grants, Fiscal Years 1980 and 1989
1989 Dollars

Item	1980	1989	Percent Change 1980–89
Federal Grants to Massachusetts (millions)	$787	$561	−29
U.S. Total (billions)	30	24	−18
Federal Grants as a Percent of Capital Outlays in Massachusetts	46	20	
U.S. Total	36	22	

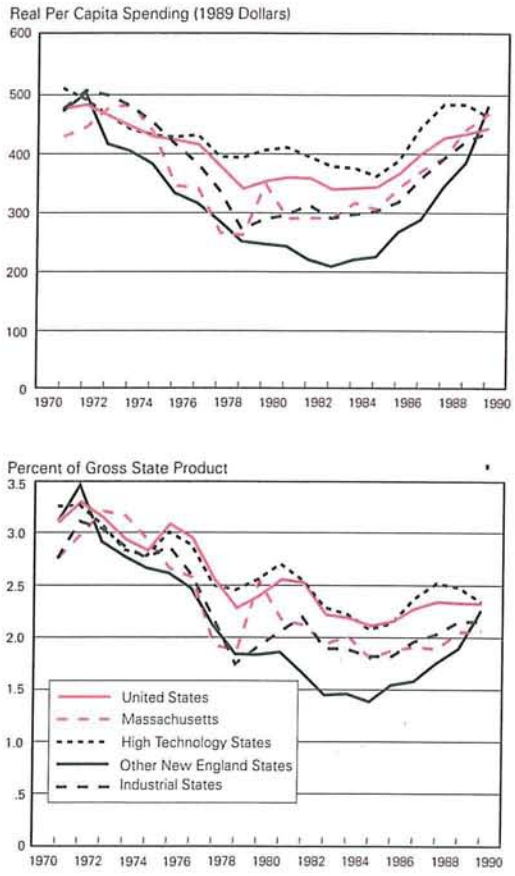
Source: Table 2 and Appendix Table B1.

On the other hand, Massachusetts appears to have been relatively hard hit by the cut in federal grants during the 1980s. Total federal money flowing to the Commonwealth in constant dollars declined from $787 million in 1980 to $561 million in 1989, or roughly 29 percent, compared to a decline of about 18 percent for the nation as a whole (Table 4). As a result, Massachusetts fell during the 1980s from enjoying a significantly greater federal contribution to capital expenditures than other states to being slightly below average. This decline largely reflects the holdup in highway money for the Central Artery Depression/Third Harbor Tunnel project. Now that funds for 80 percent of the project have been appropriated, Massachusetts is receiving enormous amounts of federal highway money. Data for fiscal 1991 show $737 million, or 14 percent of total federal highway funds, flowing to Massachusetts; this is the largest single amount going to any state and accounted for nearly half of the money allocated to the combined Middle Atlantic and New England regions.

Figure 1 compares Massachusetts state and local capital expenditures since 1970 with the average for the United States, other New England states, the

Figure 1

State and Local Capital Spending, Fiscal Years 1970 to 1989



Source: U.S. Bureau of the Census, *Government Finances*, various years; U.S. Bureau of Economic Analysis, Gross State Product, Machine Readable Data.

industrial states, and the high technology states, measured first on a per capita basis in 1989 dollars and then as a percent of Gross State Product (GSP).[5] The most striking feature is the overall pattern of spending: real per capita expenditures for infrastructure began to decline in the early 1970s, dropped sharply in the wake of the 1974–75 recession, and bottomed out—at a very low level—in the recession of 1981–82; since 1982 expenditures have been rising almost everywhere. (Individual state data appear in Appendix Table B2.) The pattern for spending as a percent of GSP is similar. (See Appendix Table B3 for state-by-state data.) Within this great swing, Massachusetts, which has consistently spent more than other New England states and less than the average for the nation or the high technology states, looks very much like the other industrial states.

The outcome of this spending on capital investment is that Massachusetts' stock of public capital, measured in constant dollars on a per capita basis, has converged toward that of the comparison states (Appendix Table B4).[6] Nevertheless, Massachusetts' per capita public capital remains slightly below that in all the comparison states except New England (Table 5). Interestingly, Massachusetts and New England as a region also have relatively low levels of private capital per person. The differences in both the public and the private capital stocks suggest that Massachusetts and New England more generally have tended to substitute highly skilled labor for physical capital.

Not only does Massachusetts have relatively low levels of capital per person, but also it no longer excels in maintenance of its infrastructure. The Federal Highway Administration regularly publishes information on the miles of pavement and number of bridges rated deficient. As shown in Table 6, almost 10 percent of highway miles in the United States are rated deficient. (See Appendix Table B5 for individual state data.) Looking solely at the 1989 data, Massachusetts appears comparable to the nation.

Comparing data from the early 1980s with data for 1989, however, reveals a disconcerting trend. While the percentage of highway mileage rated deficient has fallen for the United States as a whole and

Table 5

Public and Private Capital Stock Per Capita, 1989

States	Per Capita Stock of: Public Capital	Per Capita Stock of: Private Capital	Ratio of Public to Private Capital Stock
U.S. Total	$6,860	$18,804	.36
Massachusetts	6,598	16,510	.40
Other New England States	5,907	15,609	.38
Industrial States	7,216	17,010	.42
High Technology States	6,706	19,162	.35

Source: Authors' estimates. See Appendix Table B4.

noticeably for the industrial and high technology states, it has increased somewhat for other New England states and tripled for Massachusetts. Thus, the condition of Massachusetts' infrastructure now could be viewed as about average, but a continuation of the trends of the last decade will soon put Massachusetts (and New England) at the bottom of the heap.

The data for bridges reveal a similar pattern. For the nation as a whole, the problem is serious, with more than 40 percent of all bridges rated as deficient. Massachusetts in 1988 looks slightly better than the nation and the industrial states. But again, comparing 1980 with 1988 reveals a disturbing pattern; while the percentage rated deficient nationwide has remained relatively stable, the percentage of Massachusetts bridges rated deficient has more than doubled. The current data show that Massachusetts simply reflects the national problem of infrastructure disrepair, but the trends of the 1980s indicate that the Commonwealth's problems could soon become significantly more serious.

One area where Massachusetts' infrastructure improved was mass transit. During the 1980s, the Massachusetts Bay Transportation Authority (MBTA) undertook substantial capital investment. It expanded the Red Line northward from Harvard Square to Alewife, depressed and relocated the Orange Line, renovated South Station, and replaced rolling stock. Urban Mass Transportation Administration statistics show that from 1979 to 1988 the average age of the MBTA's fleet declined from 12 years to 11 years, mirroring the improvement nationwide. At the same time, the number of MBTA road calls per 10,000 vehicle revenue miles fell from 6.4 to 2.2, reflecting both the drop in age of the fleet and improved maintenance procedures.

Putting all of the pieces together suggests two conclusions. First, from today's perspective, the picture is neither overly encouraging nor discouraging. Massachusetts suffers from the same problems plaguing all states—declining federal dollars and increased local demands. This has produced a situation where roads and bridges are in disrepair and mandated environmental work remains undone; on the other hand, Massachusetts' major mass transit system has been improving. At this point, the condition of Massachusetts infrastructure must be categorized as average. Whether this is the position the state would choose for itself, given the importance of public capital in promoting growth, is another question.

A less sanguine conclusion emerges from the trends of the 1980s. Massachusetts appears to have received relatively little in the way of federal grants and undertook relatively little infrastructure investment, outside of the activities of the MBTA, during the 1980s. The primary reason seems to be that state officials were preoccupied with planning for the Central Artery/Harbor Tunnel project, which ran into serious opposition in Washington. Regardless of the explanation, statistics on the condition of roads and bridges show rapid deterioration in the quality of Massachusetts' infrastructure during the last ten years. Although Massachusetts is now roughly at the national average, another decade of neglect could create serious problems.

Table 6
Infrastructure Quality, Selected Years

	Percentage Rated Deficient[a]			
	Highway Mileage		Bridges	
States	1982	1989	1980	1988
U.S. Total	13.7	9.5	40.5	41.3
Massachusetts	3.2	9.8	18.8	38.7
Other New England States	11.9	12.5	37.4	46.8
Industrial States	16.2	8.3	38.0	40.9
High Technology States	13.2	7.1	36.1	33.6

[a]The Federal Highway Administration (FHA) uses a numerical rating system from 0 to 5 to quantify pavement condition. Zero represents pavements that are extremely deteriorated, perhaps needing complete reconstruction, and 5 represents pavements in very good condition, usually only new or recently resurfaced pavements. Deficient pavement carries a rating of 2 or less, or 2.5 or less for the Interstate System, which requires a higher standard of performance. A similar system is used in rating bridges. The sufficiency rating of a bridge is a weighted composite of three major areas: structural adequacy and safety (55 percent), serviceability and functional obsolescence (30 percent), and essentiality for public use (15 percent). The lower the sufficiency rating of a bridge, the higher its priority for replacement or rehabilitation. A bridge is classified as deficient if its rating is 80 or below.

Source: See Appendix Table B5.

III. Decision-Making in Massachusetts

How are physical capital investment decisions made in Massachusetts? Who decides on the tradeoffs between current and capital expenditures and between different types of investments? The most striking feature of capital spending in Massachusetts is that only a small portion of the spending plans goes through the state budget process (Table 7).[7] Instead,

Table 7
Public Capital Expenditures in Massachusetts by the State, Authorities, and Localities, 1989

Level of Government	Public Capital Expenditure (Millions)	Percent of Total
State	$ 698	25
Authorities	990	35
Massachusetts Bay Transportation Authority	346	12
Massachusetts Port Authority	53	2
Massachusetts Water Resources Authority	135	5
Massachusetts Turnpike Authority	37	1
Other	418	15
Localities	1,116	40
Total	$2,804	

Note: The figures for total state and local spending and for local spending come directly from the publication *Government Finances*. The figure for state spending is the Census Bureau figure for state spending minus Census Bureau tabulations of capital spending by off-budget entities. Local spending here represents spending by cities, counties and towns. Spending by level of government was not yet available for 1989, thus an estimate for spending of cities, counties and towns was made as follows: the share of 1988 local capital spending done by cities, counties, and towns was applied to the 1989 total of local government capital spending. The total for authorities is then the remainder. Within the authorities category, the numbers for the Massachusetts Bay Transportation Authority, the Massachusetts Water Resources Authority, the Massachusetts Port Authority and the Massachusetts Turnpike Authority are derived from the authorities' annual reports or from their Treasurer's office.

Source: U.S. Bureau of the Census, 1989, *Government Finances in 1987–88*; U.S. Bureau of the Census, 1990, *Government Finances: 1988–89 (Preliminary Report)*; U.S. Bureau of the Census, unpublished tabulations; Authority annual reports.

public authorities play the major role, creating a fragmented and uncoordinated maze of decision-making. Massachusetts is not unique in this regard; special districts and public authorities have proliferated in many states, particularly in response to recent state limitations on local property tax revenues and debt issuance.

The state government itself makes capital investments for construction and renovation of state buildings (including correctional facilities), environmental projects, housing, and transportation. (The Central Artery/Harbor Tunnel project is a state initiative.) On paper, the state appears to take a systematic approach to evaluating capital spending initiatives. The *Capital Budgeting Procedures Manual* contains detailed instructions to guide state agencies in preparing their long-range capital development plans. These plans are then transmitted to the Division of Capital Planning and Operations within the Office of Administration and Finance. They are reviewed, ranked by importance, approved by the Governor, and then sent, in the form of a separate capital budget, to the legislature. The legislature can reassess the relative merits of the various investment initiatives and compare the merits of spending on capital projects as opposed to current services.

While the approach seems reasonable, it has two problems. The first is the usual phenomenon that as soon as budget pressures emerge, capital plans get squeezed out in favor of spending on current services. As Governor Dukakis indicated in his 1991 Capital Budget recommendations, capital spending restrictions were instituted in November 1988 in order to "ensure that the Commonwealth's capital program remains affordable from year to year and that debt service costs do not crowd out other important operating budget priorities" (Commonwealth of Massachusetts 1990b, p. I-2). The spending controls limited capital spending in fiscal 1989 and 1990 and will continue to do so in the future. As a result of the controls on spending, the Commonwealth carried forward into fiscal 1991 unissued bond authorizations totaling more than $6 billion, and the Governor requested no increase in appropriations for capital projects in 1991.

The second problem is that most of the non-local capital spending decisions are not made at the state level but rather reflect a series of independent initiatives by public authorities. The concept of a public authority was imported from England and first used in the United States for constructing ports in New York and New Jersey. Robert Moses of New York embraced the concept in the late 1930s to build an innovative network of roadways, tunnels, and bridges that was the marvel of contemporary transportation. Moses characterized authorities as "non-political organizations in which the speed, flexibility and absence of red tape, traditionally associated with private industry, could be used for public purposes" (Caro 1974, p. 633). The question is whether this romantic assessment applies to Massachusetts' independent authorities today. Do they enhance or hinder the state's effort to have a rational program of infrastructure construction?

Since much of the state and local capital spending in Massachusetts is done by four major authorities (20 percent of the total), it is useful to take a quick look at why they were established, how they operate, and the nature of their relationship to the state. (For further information on the authorities, see the first box.)

Massachusetts' Public Authorities

The Commonwealth currently has over 500 individual authorities. Although the exact structure of the authority varies with the project's financing needs, the roles to be played by existing subdivisions, and numerous other administrative considerations, a general description is possible. An authority is a public entity established by the state legislature to perform specific tasks that have a public purpose. A board of directors, appointed by the Governor, is responsible for the entity's activities, and the board appoints a director to administer the authority. Employees of an authority are not subject to civil service or other state personnel laws. Authorities are provided with the means to fund their activities, and therefore can operate independently of the state budget process. They can generally issue debt to finance capital expenditures and that debt is free of state and

Box Table 1
Major Public Authorities in Massachusetts, 1989

Name	Year Established	Number of Employees	Operating Budget (million $)	Capital Budget (million $)
Operating Entities				
Mass. Bay Transportation Authority	1964	6,710	629.6	346.1
Mass. Municipal Wholesale Electric Co.	1975	140	205.0	
Mass. Water Resources Authority	1985	1,712	182.9	135.1
Mass. Port Authority	1956	968	119.0	53.1
Mass. Turnpike Authority	1952	1,398	117.0	37.3
Steamship Authority	1960	487	25.2	
Mass. Convention Center Authority	1982	101	14.6	12.4
Mass. Technology Park Development Corporation	1982	45	2.7	[a]
Bay State Skills Corporation	1981	18	2.3	[a]
Mass. Corporation for Educational Telecommunications	1982	7	1.2	
Government Land Bank	1975	15	1.0	[a]
Mass. Technology Development Corporation	1978	10	.8	[a]
Community Economic Development Assistance Corporation	1978	7	.6	n.a.
Community Development Finance Corporation	1975	8	n.a.	[a]
Worcester Business Development Corporation	n.a.	n.a.	n.a.	n.a.
Financing Entities				
Mass. Housing Finance Agency	1966	236	15.2	—
Mass. Educational Loan Authority	1981	13	12.1	—
Boston Metropolitan District	1929	1	11.3	—
U. Mass. Building Authority	1963	0	15.7	—
U. of Lowell Building Authority	1961	25	2.7	—
State College Building Authority	1963	4	.7	—
Southeastern Mass. U. Building Authority	1964	0	.1	—
Mass. Industrial Finance Agency	1978	23	3.2	—
Mass. Health & Educational Facilities Authority	1968	19	1.8	—
Total 24 Authorities		11,947	1,366	613.4
State		93,550	12,641	698.1

n.a. = not available.

[a]According to the *Annual Financial Report*, these five authorities together in 1989 spent $29.4 million for acquisition of fixed assets.

Source: "Massachusetts Public Authorities," 1990, Background Report to Crozier Commission; U.S. Bureau of the Census, unpublished tabulations; Commonwealth of Massachusetts, Office of the Comptroller, 1990, *Comprehensive Annual Financial Report for Fiscal Year 1989;* Authority Annual Reports.

federal tax.

Authorities can be classified as either financing or operational entities. Most financing entities issue tax-exempt bonds to reduce the cost of financing certain public purpose activities, such as affordable housing (Massachusetts Housing Finance Agency) or tuition loans (Massachusetts Educational Loan Authority). Operational authorities provide many basic public services to the residents of the Commonwealth and are the ones responsible for the capital investment.

Twenty-four authorities account for the bulk (roughly 90 percent) of authority employment and spending. In 1989, this group employed 12,000, compared to total state employment of 94,000, and had operating budgets totalling $1.4 billion, compared to the state figure of $12.6 billion (Box Table 1). Spending for capital projects is even more concentrated and larger relative to the state. Four authorities, the MBTA, MWRA, Massport, and the Turnpike Authority, spent almost as much on capital projects in 1989 as the entire state government.

This pattern was true for the entire 1980s (Box Table 2). The big spender was the MBTA, which undertook major capital improvements. Two of the largest projects involved expanding the Red Line north from Harvard Square to Alewife and relocating and depressing the southern portion of the Orange Line. The MBTA also bought new rolling stock and upgraded other tracks and platforms.

Lesser amounts were spent by the other three authorities. The Turnpike Authority repaired bridge decks, resurfaced roadways, and improved the tunnels. Since its creation in 1985, the MWRA has been replacing water pipes and improving its capacity to handle sewer overflows. Massport completed projects at Logan Airport, including a new international terminal and soundproofing schools and homes in the area, upgraded facilities at the port, and developed other waterfront property, such as the World Trade Center. It also made improvements to the Tobin Bridge, including a pipe to provide water pressure for fire fighting, a traffic monitoring system, and road and deck repairs.

While most authorities are financially independent, several receive state support in the form of operating or debt service assistance. In 1989, the state provided $353 million to the authorities, 80 percent of which went to the MBTA. The MBTA has a unique and complex relationship to the state and over the years has become an increasing drain on state resources. As indicated in the text, the legislation enabling the MBTA to issue debt also provides for state assistance toward debt service and operating costs. The state has contracted to pay 90 percent of the debt service on up to $2 billion of bonds and is responsible for the annual MBTA deficit (operating deficit plus debt service costs).

The timing of the reimbursement process for the MBTA further complicates the relationship. The MBTA operates on a calendar year basis and will cover its 1990 operating costs by issuing short-term notes (guaranteed by the Commonwealth). In December 1990, the MBTA will submit a bill to the Commonwealth, which will be in the middle of its 1991 fiscal year. The Commonwealth will bill the cities and towns for their share and put its share in the state's 1992 budget. The state receives payments from the cities and towns and reimburses

Box Table 2
Capital Spending of the State Government and Authorities, Fiscal Years 1980 to 1989
Millions of 1989 Dollars

		Authorities			
Year	State	MWRA	Massport	MBTA	MTA
1980	404	0	41	397	13
1981	503	0	64	422	11
1982	492	0	49	399	16
1983	579	0	49	402	15
1984	452	0	33	398	19
1985	513	0	61	390	32
1986	506	25	80	315	32
1987	682	47	63	299	32
1988	672	124	48	354	37
1989	698	135	53	346	37
Total	5500	332	541	3722	245

Note: The state total shown here is different from that shown in Table 8 because of differences in the methodology used to count capital spending by the state and the Census Bureau. The figures here represent Census tabulations of state spending that exclude off-budget entities. See footnote 7 for a reconciliation of Census Bureau capital spending figures with those of the state government. Numbers may not add to totals because of rounding.

Source: U.S. Bureau of the Census, special tabulations; authority annual reports; background data to Crozier Commission.

Massachusetts' Public Authorities, continued

the MBTA at the end of the Commonwealth's fiscal year 1992. Thus, a dollar spent by the MBTA in January 1990 is financed at short-term rates until the authority is paid off in June 1992.[8]

State costs for the MBTA have increased rapidly as income from fares has grown more slowly than expenses and the localities' share has declined in the wake of Proposition 2½. State assistance represented 59 percent of total 1989 MBTA expenses, compared to 41 percent in 1980. Despite this large contribution, the state has little authority over MBTA operating or capital expenditures.

In addition to operating or debt service assistance, several authorities also have a state debt guarantee. The debt of local housing authorities and higher education building authorities is backed by the Commonwealth's full faith and credit. Bonds of the MBTA, regional transit authorities, the Convention Center Authority, and the Steamship Authority represent contingent liabilities of the state, which means that the state must provide the authorities with sufficient funds to meet interest and principal payments if the funds are otherwise not available. Even if the state had no explicit commitment to authorities' bondholders, the need to protect its bond rating would probably force the Commonwealth to assist any authority in financial difficulty.

On paper it may appear that the state has some knowledge and control over authority activities. The Secretary of Environmental Affairs is the Chairman of the MWRA's Board, the Secretary of Transportation is the Chairman of the MBTA'S board, and the Governor appoints board members of the other authorities. However, no formal centralized process exists for assessing the relative merits of their plans or even for collecting employment and financial data. The MBTA and the MWRA have Advisory Boards, which must approve their budgets, but the Boards have no way to make trade-offs among authorities or between authorities and state agencies. Furthermore, whatever limited information emerges is usually distributed only to the executive branch; the legislature and the public are rarely informed.

The Senate Ways and Means Committee (1985) noted that 1979 legislation that required all authorities with bonding power (except Massport and the Turnpike Authority) to file quarterly reports, detailing their outstanding and unissued bonds, projected debt service over the next two years, and bonds to be sold over the ensuing year, has been virtually ignored. In short, authorities represent a significant financial responsibility of the state, but the state has almost no control over their activities.

The Massachusetts Turnpike Authority was established in 1952 and charged with the construction, operation, and maintenance of the Massachusetts Turnpike, a 135-mile toll highway running from the western border of Massachusetts to the City of Boston. Why an authority? The need for such a road was indisputable; the difficulty was that the state had already committed unprecedented sums to highway bond issues. Thus, the Massachusetts Turnpike Authority was born out of the demand for a critical public works project in a time of diminished revenues. The idea of a public authority was not untested. Moses had used them in New York, and Maine and New Jersey had just created authorities to manage the construction of their high-speed toll roads. Authorities were very popular throughout the country in the early 1950s.

In 1958, the legislature authorized the Turnpike Authority to construct the Callahan Tunnel and to operate and maintain both the Sumner and Callahan Tunnels, one-mile harbor crossings connecting Boston with East Boston and Logan Airport. All funds for the maintenance, capital improvement, operation and policing of these facilities, as well as payment of principal and interest on bonds issued, are derived solely from tolls and other revenues generated by users. In other words, the Authority is entirely self-supporting and receives no money or guarantees from the state. It is directed by a board of three members appointed by the Governor, one of whom is designated as Chairman.

The Massachusetts Turnpike Authority was scheduled to dissolve in 1992 upon repayment of the original bonds. Instead, the Turnpike has a $603 million capital plan to rebuild the road and a $58 million program of repair and reconstruction for the tunnels. The Authority estimates that these projects will require doubling its current annual capital expen-

diture. In anticipation, the Turnpike Authority has already raised tolls on both the turnpike and the tunnels, and has shown no evidence of planning ever to dissolve itself.

In 1956, the legislature created the Massachusetts Port Authority (Massport) as an independent authority charged with the operation, maintenance, and improvement of the Tobin Bridge, the seaport, and most important, Logan Airport. The motivation for establishing a separate entity was the belief that the vital links in the Commonwealth's transportation system were so critical to the well-being of Massachusetts that they needed to be overseen by a single entity, which would not only manage their day-to-day operations but also plan for the long term. The hope was that such a structure would protect these projects from changes in the economic and political climate. The Massport Board consists of seven members appointed by the Governor for staggered terms of seven years each. It is financed by user charges and debt, and has an extensive program of maintenance, improvements, and new construction.

Four major authorities now undertake almost as much capital spending each year as the entire state government.

The Massachusetts Bay Transportation Authority (MBTA), which was established in 1964, was charged with developing, financing, and operating mass transportation for the 78 cities and towns within its jurisdiction.[9] The MBTA has a complex financial relationship with the state. Technically, the MBTA has its own bonding authority and the bonds are not backed by the full faith and credit of the Commonwealth. On the other hand, the state by statute has contracted to pay 90 percent of the debt service on up to $2 billion of bonds. (The MBTA currently has $1.1 billion outstanding.) In addition, the Commonwealth is responsible for the annual MBTA deficit (operating deficit plus debt service costs). The state can assess the cities and towns for a portion of the cost, but, in the wake of Proposition 2½, it provides most of the subsidy.

In 1985, the state legislature created the Massachusetts Water Resources Authority (MWRA) in response to a 1984 warning from Judge Paul Garrity and a pending lawsuit. Garrity had threatened to halt all new sewer hook-ups in the City of Boston, because he believed that not enough was being done by the Commonwealth to clean up Boston Harbor—among the nation's filthiest waterways—and new sewer hook-ups would lead to even worse pollution.

Responsibility for water and sewer systems at that time rested with the Metropolitan District Commission (MDC), whose budget came under the control of the state legislature. Because the MDC was perpetually underfunded, the water and sewer system serving almost half of the state's population had fallen into terrible disrepair. Aqueducts leaked millions of gallons of clean water; rain caused raw sewage to overflow from old sewer pipes into local rivers and into Boston Harbor; and the sewage that did reach the main treatment plant on Deer Island received minimal treatment and was dumped into the Harbor.

The Massachusetts Water Resources Authority took over from the MDC all responsibility for improving the water quality of Boston Harbor and modernizing the vast water and sewer system. The Authority is governed by an eleven-member board, whose chairman is the Secretary of the Office of Environmental Affairs. It is financed by user charges and debt, and receives no support from the state in the form of revenues or guarantees. Some observers suggest, however, that residents may find the scheduled doubling (in real terms) of water bills over the next ten years to pay for the cleanup of Boston Harbor unacceptable, and may force the MWRA to come to the legislature for financing in the future (Commonwealth of Massachusetts, Senate Committee on Ways and Means 1989).

These four authorities, the Turnpike Authority, the MBTA, Massport, and the MWRA, now undertake almost as much capital spending each year as the entire state government (Table 7). The desirability of such an arrangement is not a simple issue. On the one hand, the authorities were created for good reasons. Their structure permits the development of a focused and politically insulated organization that is not subject to the vagaries of the annual budget process or the rigidities of civil service or contract bidding laws. It is an attractive option when projects require the bridging of existing government jurisdictions, or when the state wants to ensure that individuals living in, say, a huge urban area of a largely rural state pay for significant public capital investments for which they are the primary beneficiaries. Financial

advantages arise from the taxpayers' willingness to pay user charges for specific services and from underwriters' preferences for a dedicated stream of revenues outside of the Commonwealth's general funds.

Moreover, the authorities appear to have had considerable success in accomplishing their assigned tasks. The underfunding of the old MDC emphasizes the problems with raising the money for necessary capital expenditures through the state budget process. The limits established in 1988 on state government capital expenditures further demonstrate the vulnerability of capital projects in periods of serious budget constraints.

On the other hand, the authorities fragment decision-making enormously, since each entity comes up with its own plans for capital projects and funds them out of its own financial sources. No process exists for evaluating the merits of providing an additional toll collection lane on the Mass Pike as compared to a new trolley car for the MBTA, or for comparing the desirability of capital spending by the authorities in general with the state's need for new prisons and other facilities.

In addition to the lack of coordination, the authorities are in the privileged position of controlling a monopoly with no regulation of their activities. The steady flow of income from fees and charges creates little incentive to search for least-cost approaches to solving projects or to forgo activities with relatively small returns. Newspaper reports and opinion polls indicate that residents of the state believe that, in some cases, the authorities have abused their positions.

Two of the authorities, the Massachusetts Turnpike Authority and Massport, both appear flush with revenues; most of the infrastructure under the control of these authorities also appears in excellent condition.[10] One could question, from the overall perspective of the state, whether an additional capital expenditure for, say, the turnpike represents the best use of public money. If it does not, the issue becomes how to reallocate the funds.

Dissolving authorities once they have completed their missions is certainly one option; tolls or other charges could then accrue directly to the state. These monies could be placed in a special fund allocated to public capital investment.

An alternative is to retain the organizational structure and expand the authorities' responsibilities; this would relieve some of the budget pressure on the state. Specific suggestions made by the Massachusetts Senate Ways and Means Committee include: Massport assuming the financing of certain tourism and economic development programs currently funded in the Executive Office of Economic Affairs; and the Massachusetts Turnpike Authority assuming the cost of snow and ice removal operations for the entire state highway system, currently paid by the Department of Public Works.[11]

Another variant is to transfer major capital projects, rather than operating activities, to the exist-

Table 8
Capital Spending by the State and Four Major Authorities, 1980 to 1989 and 1990 to 1999
Billions of 1989 Dollars

	1980 to 1989			1990 to 1999		
Entity	Total	Federal	State/ Local	Total	Federal	State/ Local
State	7.3	2.3	5.0	17.1	7.2	9.9
Authorities						
Massachusetts Bay Transportation Authority	3.7	2.0	1.7	3.6	1.6	2.1
Massachusetts Port Authority	.5	.1	.5	2.2	.1	2.1
Massachusetts Water Resources Authority	.3	.1	.2	4.5	.2	4.3
Massachusetts Turnpike Authority	.2	0	.2	.6	0	.6
Total	12.2	4.5	7.7	28.1	9.1	19.0

Note: Columns may not sum to totals due to rounding.
Source: See Appendix Table B6.

ing authorities. Such a shift would provide a source of funding and steady administration in a rapidly changing political and economic environment. One obvious option is to make the Turnpike Authority responsible for the Central Artery/Harbor Tunnel project, or merge Massport and the Mass Turnpike Authority and place the new entity in charge of the scheduled effort.

The proposed increases in responsibility, however, do not resolve the fragmented nature of decision-making for capital expenditures. On efficiency grounds, it would be better to integrate all capital spending plans in order to establish the relative importance of the individual projects. This would require standardization of reporting, which does not currently exist. It would also require some way of trading off the proposals of the self-supporting authorities with those that must be financed through the general fund. This may be difficult, but it seems desirable given that the four major authorities plan to spend almost $11 billion on capital projects over the next 10 years (Table 8). The difficulty, of course, is that many more parties become involved in the decision-making, which may complicate and delay the process. Skeptics of centralizing characterize the choice as one between disaggregation and professionalism on the one hand, and comprehensiveness and political wheeling and dealing on the other.

IV. Massachusetts' Capital Spending Plans for the 1990s

Massachusetts' capital spending plans for the 1990s consist of three pieces: activities of the state government, the mega projects, and initiatives of the authorities. In each case, the money slated to be spent in the coming decade dwarfs expenditures during the 1980s.

The State's Plans

According to the Census Bureau, the Commonwealth of Massachusetts (excluding the authorities) did not spend a great deal on capital projects for most of the 1980s. In constant dollars, annual outlays hovered between $400 million and $500 million and generally declined as a share of total state spending. It appears that, around 1986, government officials began to recognize that important infrastructure projects had been deferred. In part, this may have been a response to the work of the National Council on Public Works Improvement, which highlighted the nationwide problem of infrastructure deterioration.

In Massachusetts, a Special Commission on Infrastructure Finance was established in 1986 to address the status of the state's public capital. In a 1989 report entitled *A Survey of Massachusetts Infrastructure Needs in the 1990s*, the Commission identified a host of capital spending projects. More recently, the Governor's Management Task Force completed an updated survey and found $31 billion of needed capital investment, excluding the Central Artery/Harbor Tunnel project and initiatives by the authorities. Of the total, $7 billion would be provided by the federal government, leaving $24 billion to be paid from the state coffers. The bulk of this money would go to badly needed road and bridge repairs, wastewater treatment, solid and hazardous waste disposal, state hospitals, prisons, and public housing. One possible scenario for the 1990s, then, is the state spending $2.4 billion annually to eliminate its infrastructure deficit.

Such a level of annual expenditure seems unrealistic. First, $2.4 billion is more than double the state's current outlay for capital expenditure; an increase of that magnitude is unlikely given the continuing budget problems. Second, the state has instituted a cap of $925 million on its own-source spending for capital projects, which it appears committed to meeting for the foreseeable future. The cap is in nominal dollars, but the assumption for this study is that the legislature will index the amount.

Assuming that the $925 million cap holds for the entire decade implies total state expenditures during the 1990s of $9.25 billion. On the one hand, this number looks large; total state spending during the 1980s amounted to about $5 billion. Thus, even if the limit is adhered to, capital spending in the 1990s will place increased demands on state revenues. On the other hand, $9.25 billion meets less than 40 percent of the $24 billion of infrastructure needs documented by the Governor's Management Task Force. Given the disparity between estimated needs and realistic spending levels, state officials and legislators will have to consider carefully the merits of alternative projects identified by the Commission and establish clear priorities.

The Mega Projects

The second piece of Massachusetts' capital spending plans is the mega projects. The Central Artery Depression/Third Harbor Tunnel project and the Boston Harbor Cleanup are the two largest capital

Table 9
Capital Spending on the Central Artery/Third Harbor Tunnel and Boston Harbor Cleanup Projects, Fiscal Years 1990 to 2000
Millions of 1989 Dollars

	Central Artery			Harbor Cleanup		
Year	Total	Federal	State	Total	Federal	Local
Total	4972	4325	648	2659	118	2541
1990 or before	32	28	3	181	18	162
1991	303	273	30	147	15	133
1992	455	409	45	366	19	347
1993	557	481	76	431	15	416
1994	712	615	97	479	9	469
1995	879	759	119	384	9	375
1996	841	727	114	205	10	195
1997	662	572	90	119	6	113
1998	430	371	58	153	7	146
1999	104	90	14	140	7	133
2000	—	—	—	55	3	53

Addendum: Massachusetts Capital Spending for Highways and Wastewater Treatment, Fiscal Year 1989

	Highways			Wastewater Treatment		
	Total	Federal	State & Local	Total	Federal	State & Local
1989	613	234	379	340	92	248

Note: Columns may not sum to totals due to rounding.
Source: Yearly construction expenditures and grants obtained from Central Artery/Harbor Tunnel Public Information Office and MWRA Budget Office; Addendum information from Table 3 and Appendix Table B1.

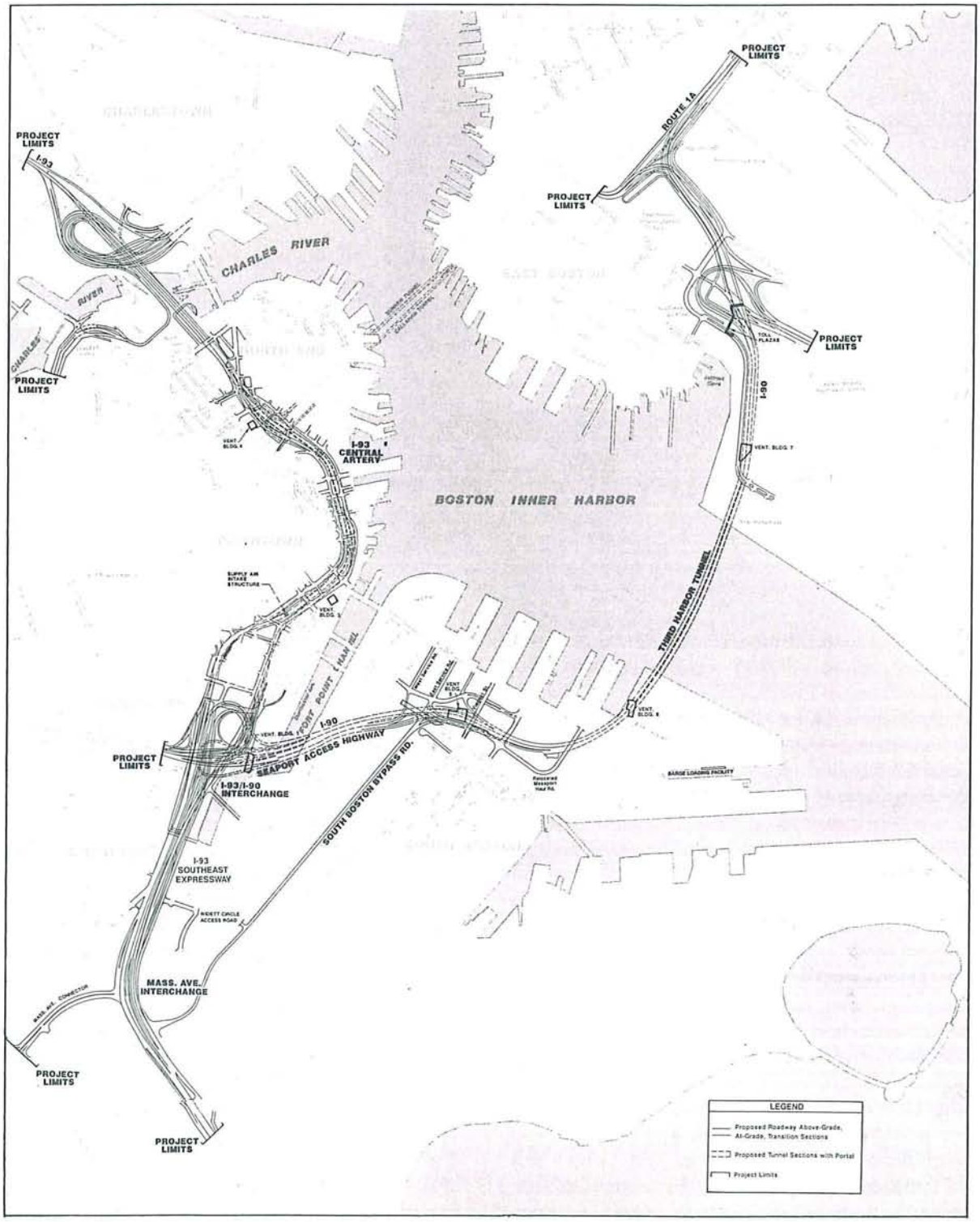
spending initiatives in the state's history. The costs of the two projects are estimated (in 1989 dollars) at $5.0 billion and $2.7 billion, respectively (Table 9).[12] The Central Artery/Tunnel is a 7.5 mile interstate highway project covering portions of both I93 (north-south) and I90 (east-west), which is currently scheduled for completion in 1999 after a 10-year construction period (Figure 2). The primary justification for this project is that, when built in the 1950s, the artery was never designed to meet interstate standards: it has too many entrances and exits too closely spaced, too much forced crossing of traffic, no breakdown lanes, and no entrance and exit lanes. Furthermore, the road was designed to carry 75,000 vehicles daily, and the volume has grown to about 190,000 vehicles daily. Even if it were not rebuilt entirely, the existing structure needs substantial repair. The repair would entail significant traffic disruption and would not produce the capacity, safety, and environmental improvements included in the new design.

The project was made eligible for federal Highway Trust Fund money in the 1987 Surface Transportation Act. Currently 80 percent of the $5.0 billion (1989 dollars) project has been approved for interstate completion funds; the funding is structured to cover inflation's effects on project costs and will be available until the project is completed. Under the current interstate program, the federal government contributes 90 percent of the funds required for new construction on the interstate network, and the most likely scenario is that the entire portion of the approved 80 percent will be eligible for a 90-percent federal match.

A very slight possibility exists that the match on the approved portion might turn out to be somewhat lower. The interstate highway system is now largely complete and the Bush Administration wants to revamp the highway grant program when Congress reauthorizes the Surface Transportation Act in 1992. Two key components of the restructuring would be a shift from narrow categorical programs to broad multi-purpose grants and an increase in the share of total costs paid by the states. The Administration proposes that the federal matching share for new

Figure 2

Map of the Central Artery/Third Harbor Tunnel Project



Source: Central Artery/Third Harbor Tunnel Public Information Packet.

grants be reduced to 75 percent. If for some reason the Central Artery/Tunnel project were not grandfathered, money received after 1992 could be subject to a lower federal match. Such an outcome is unlikely, but not impossible.

The other 20 percent of the project, the portion of the Artery between High Street and North Station, is currently eligible for 90 percent federal funding through the "4R" program (for resurfacing, restoring, rehabilitating, and reconstructing interstate state roads) or other federal programs, but funding for this portion of the program has yet to be authorized. Congress is expected to decide on funding for the High Street–North Station segment as part of the reauthorization of the Surface Transportation Act in 1992. Proponents of the project say that no approved interstate highway project in the history of the system has ever been left incomplete due to lack of funds, and see large amounts of money available for refurbishing now that construction of the interstate highway system is virtually finished.

On the other hand, if the Bush Administration proposals are adopted, it is quite possible that no special money would be available for a project such as depression of the High Street–North Station portion of the Artery. Instead, the state would receive a multi-purpose highway grant for the Commonwealth's major roads and would have to decide whether to spend the money on the Artery or other desirable highway programs. Moreover, the new grant money would be unlikely to cover 90 percent of the costs; as noted earlier, the percentage will probably be 75 percent.

One additional element further complicates the picture. Even if the entire project becomes eligible for federal funding, the annual costs must first be financed by the state and then reimbursed by the federal government. Thus, the feasibility of raising the initial money becomes an issue, given the state's current fiscal situation. Uncertainty about the state's ability to carry out its share of the financing could make the federal government wary about committing or recommitting resources to the project, particularly given the hostility in Washington to this project.

The state has taken a step to demonstrate its commitment to the Central Artery/Tunnel project. The legislature raised the gas tax by 6 cents per gallon effective July 1990, and by another 4 cents per gallon beginning January 1991, nearly doubling the previous 11 cents per gallon tax. The bulk of this revenue increase will be deposited in an infrastructure fund, an account within the state's highway fund. The infrastructure fund will be used to pay debt service on special obligation revenue bonds for infrastructure projects, including the Central Artery.

The same legislation also says that "not more than ten percent" (Section 93 of Chapter 121 of the Acts of 1990) of *all* gas tax receipts can be devoted to the Central Artery Project. If the state spends less than 10 percent in any year, it can carry over the unexpended balance to any future year provided that the sum of expenditures from the carryover and the current year do not exceed 20 percent of gas taxes collected in that fiscal year. No limitations apply to spending on the Central Artery project out of other Highway Fund receipts.

Despite this sign of commitment, the extent of federal participation has yet to be finalized. This means that the state's liability remains undefined. Under the most likely scenario, the entire approved 80 percent of the project would be eligible for the 90-percent match and the state would allocate some of the general highway money received after 1992 for depression of the Artery with a matching rate of 75 percent. This would require a contribution from the state of roughly $650 million. (A Lazard Frères and Co. report (1990) also anticipates federal funding of roughly the same magnitude, $697 million.)[13] This figure is shown in Table 9. In the worst case, federal funds for the approved 80 percent received after 1992 would be subject to the new lower matching rate and the 20 percent of the project not yet approved would receive no federal funds. In this case, the state's costs would be about $2 billion.

Given these two scenarios of federal funding, it is possible to generate rough estimates of the annual debt service requirements of the Central Artery/Tunnel project. Comparing these estimates to the constraints in the 1990 legislation can shed some light on the feasibility of the state's funding plans. Assuming an 8 percent coupon rate, the state could support annual debt service in the most likely case of federal funding. In the worst case, however, the state could not support debt service solely from gas tax receipts, even including the carryover provisions. The state would have to draw on other highway fund receipts in the latter years of the project to cover annual debt service costs. While this is not impossible, it would be difficult given the competing demands on highway fund resources from other highway projects and operating expenses.

While the Central Artery/Tunnel project is a state initiative financed by federal funds, the Harbor Cleanup is a federally mandated project that must

rely primarily on local revenues. In July 1985, Federal Judge A. David Mazzone found that Massachusetts had violated federal water pollution control laws.[14] In May 1986, he issued an order and a schedule for the cleanup of Boston Harbor. By December 1991, sludge disposal into Boston Harbor must stop, and by 1994 one-half of the construction of a new primary treatment plant must be completed, with the entire plant operational by 1995. Construction on a secondary treatment plant must start in 1992, with one-quarter of the plant completed by 1996, and the entire plant operational by 1999. These two plants, when completed, will be the second largest system in the country, able to treat a peak flow of 1.2 billion gallons of sewage daily.

The Harbor Cleanup project is expected to cost approximately $2.7 billion in real (1989) dollars over the construction period, which began in 1988 with completion now scheduled for 2000. Limited federal help is available, but the project will be financed primarily through bonds, which will be paid off by increasing user charges. The Massachusetts Water Resources Authority (MWRA) estimates that the average homeowner in the service area will see a rise in the water and sewer bill from the current level of $350 annually to more than $600 (in 1989 dollars) by the year 2000.

Despite the costs, most reports indicate that residents of Massachusetts support the mega projects, particularly now that other construction activity has plummeted and the rest of the economy has weakened. In addition to remedying serious infrastructure needs and improving the environment, the projects will boost local employment and income. At their peak these two projects could generate on the order of 25,000 jobs.[15] In the average year, however, the employment gains are substantially lower at approximately 13,000 jobs (8,900 from the Artery and 4,100 from the Harbor Cleanup), which represents less than 1 percent of total employment in the Greater Boston area in 1988. Given the current economic climate these jobs will certainly be beneficial, but should not be expected to reverse the current slowdown.

On the other hand, if both the Central Artery/Tunnel project and Harbor Cleanup proceed as scheduled, the capital spending picture and the contribution of the federal government to these efforts will look very different in the 1990s than the 1980s. Spending for the Central Artery/Tunnel project will amount to $879 million in 1995, nearly twice the 1989 level of spending for highways; the annual federal contribution (under the most likely scenario) far exceeds any highway grants received by Massachusetts in recent decades (Table 9). The annual spending for the Harbor Cleanup alone during the 1990s will roughly equal all spending by the state and localities on wastewater treatment in 1989. The vast majority of the financing for the Harbor Cleanup will be provided locally, with a small amount of state support available in the form of loans from the State Revolving Fund, so the project does not change the grant situation in this area. Nevertheless, if both projects are completed, Massachusetts in the year 2000 should look like a state that undertook major investment in infrastructure and received considerable federal support.

Authority Initiatives

The spending plans of the independent authorities constitute the final piece of Massachusetts' capital spending picture for the 1990s. Massport, the Turnpike Authority, and the MBTA all have significant initiatives for the coming decade (Table 8). The MWRA also has capital plans in addition to those projects directly related to the Harbor Cleanup.

Massport's investment program for 1990 to 1999 calls for $2.2 billion dollars of improvements, including terminal and runway enhancements and noise abatement efforts at Logan Airport, port improvements, and major rehabilitation on the deck and supporting piers of the Tobin Bridge.

The Turnpike Authority plans to reconstruct many of the bridges along the turnpike, resurface about 10 miles of roadway annually, and replace one-quarter of the guard rail which is over 30 years old, at a total cost of $603 million. This total planned spending also includes upgrading toll plazas to reduce congestion and improving service and maintenance areas. The Authority also plans a $58 million dollar program of repairs to the Sumner and Callahan tunnels, including ceiling and tile replacement, renovations of ventilation systems, and roadway resurfacing.

According to the Governor's Management Task Force, the MBTA needs to spend about $4 billion over the next ten years to purchase rolling stock and upgrade tracks and stations.

In addition to the Harbor Cleanup already discussed, the MWRA plans expenditures of about $2 billion during the next ten years. Its capital plan includes, on the water supply side, replacing corroded water pipes, rehabilitating pump stations, improving chlorination facilities, and upgrading aque-

ducts. Similar improvements are planned for the sewer system, primarily replacing old pipes and pump stations.

Summary

It is apparent that Massachusetts as a whole has ambitious capital spending plans for the 1990s. While these estimates are uncertain, especially in the later years, they represent a monumental increase in capital spending over the previous decade (Table 8). The estimate of state spending in the 1990s is almost double its spending during the 1980s. MWRA, Massport and the Turnpike Authority all have plans that represent even larger increases over the 1980s. The current estimate for the MBTA is about the same as its spending over the previous decade.

If the state and the authorities actually undertake this ambitious plan, Massachusetts in the year 2000 ought to be a state with a substantial, well-maintained infrastructure. Whether this plan is realistic is another question. Massport and the Turnpike Authority appear to be in sound financial condition and thus able to carry out their plans. The MWRA currently has no leeway where the Harbor Cleanup is concerned, so this project must proceed. The MWRA also has a good credit rating, which puts it in a favorable position to implement its other programs. However, if rates increase too quickly, it may run into resistance from customers and be forced to curtail its efforts. The MBTA's projected expenditures, which are about equal to 1980s levels, may well be curtailed by the state's fiscal problems. While the state itself needs to undertake capital investment, it may also face substantial difficulties unless it can bring its operating budget under control.

V. Conclusion

Massachusetts' infrastructure, like public capital in the rest of the country, is in need of repair. Moreover, Massachusetts faces the huge task of cleaning up Boston Harbor. The Commonwealth's plight is not difficult to understand; the federal government has reduced grants to states and localities, while localities are faced with mandated environmental projects and caps on their property taxes. As a result, the demands on state government resources have expanded enormously, and capital spending is always the first item to go during periods of budget pressures.

To compensate for low spending during the 1980s, to complete the Central Artery Depression/Third Harbor Tunnel project, and to execute the mandated cleanup of Boston Harbor, Massachusetts has constructed an ambitious capital spending agenda for the 1990s. A conservative estimate of the total cost of state, authority and mega project activity over the next ten years is $28 billion (in 1989 dollars), $19 billion of which must be financed either directly or indirectly by the taxpayers of Massachusetts. These projections dwarf actual capital expenditures during the 1980s, which totalled $12 billion (in 1989 dollars) with $8 billion paid by Massachusetts taxpayers. Some of the funding may be derived from fees and charges, but a significant burden may fall to the state government. Hence, any long-range planning effort should incorporate the debt service costs from these capital spending initiatives in expenditure projections.

The ambitious agenda also underlines the need for coordinating efforts and establishing priorities. The state cannot control the artificial incentives created by the high matching rates of federal government grants; the only sensible response is to take full advantage of the offer. It can and must, however, eliminate the incentives for low-priority investments created by overfunded authorities. It must also carefully evaluate authority initiatives as compared to state-funded projects, and the merits of alternative state capital spending proposals. Massachusetts must create a mechanism for the oversight of the capital investment activities of both the Commonwealth and the authorities.

In this regard, Massachusetts does not have to reinvent the wheel; other states have faced the same problems and have developed a variety of approaches to make systematic assessments of their capital spending proposals. For example, in 1975 New Jersey instituted a Capital Planning Commission, which is considered by the National Conference of State Legislatures to be one of the best in the nation. Kentucky recently adopted the New Jersey model in an effort to revamp its capital planning and budgeting process. The major advantage of the New Jersey/Kentucky approach is that, through the Commission, the executive branch, legislators, and private individuals are fully involved in the capital planning process; they have consistent, detailed information on the capital plans, and the Governor has knowledge of and veto power over the activities of authorities. (See the second box.) The New Jersey/Kentucky system, or a system that at least gives the Governor information about and veto power over authority projects, is badly needed for Massachusetts.

The New Jersey Capital Budget Process

New Jersey's Capital Planning Commission was established in 1975 in response to complaints from Wall Street and bond raters that New Jersey's capital budgeting process was haphazard with no coherent methodology.

The Commission consists of four legislators (two from each house, two from each party), who have traditionally been the ranking people on their respective finance or appropriations committees; four members appointed by the Governor, who have traditionally been the Director of the Office of Management and Budget, the Treasurer, the Counsel to the Governor, and the Director of Planning and Management; and four private sector members (two from each party), who are appointed by the administration and confirmed by the Senate. The Commission has a small staff to do most of the technical work.

The annual capital planning process begins with the capital planning officer within each state agency compiling and establishing priorities for all capital requests. (The current year requests are always proposed within the context of longer range, three-year and seven-year, plans.) The officers forward the documentation to the Commission, which holds informal meetings with staff-level people who submitted the capital plans. At these meetings, the Commission members ask for justification for projects, and may request that an agency scale down its requests. Generally, the Commission tries to get the agencies to make as realistic a request as possible, given its members' knowledge of the budget process and the demands on the operating side of the budget.

The next step is a formal hearing, where the Secretary of the agency makes a personal appearance before the Commission either to review individual requests or to describe the agency's overall agenda. In the wake of the hearings, the Commission compiles a list of recommendations, which forms the basis for the Governor's capital budget.

While this process does not explicitly include the authorities, it does allow the Governor some power over authority activity. The executive branch has an office (the Governor's Authorities Unit) to review the plans and budgets of all state-level authorities, and the Governor can ask for clarification and further information on any project. The Governor has the power to veto the minutes of any independent authority and thereby stop a proposed project, if necessary.

New Jersey has also established a coordinating committee of all agencies and authorities that are involved with any aspect of transportation so that their plans can be designed to best meet the needs of the state. The chairman of the committee is the Secretary of Transportation.

Thus, the key elements of New Jersey's plan that could benefit Massachusetts are early involvement of legislators in the capital planning process, detailed and consistent information about state projects and their alternatives, and knowledge of and veto power over the activities of authorities. In serving on a capital planning commission, legislators can evaluate the merits of specific capital requests from the perspectives of individuals who see the full spectrum of the state's obligations, both operating and capital. Once they have participated in the construction of a proposed capital budget, they, as ranking members of the finance and appropriations committees, can convince their committees and other legislators that the plan is realistic and valuable. The second important element, full and consistent information, allows the Commission to make rational choices between projects. Finally, the oversight of authorities is crucial when they constitute such a large part of capital spending.

Two factors attest to the success of the New Jersey plan. First, since the establishment of the Commission voters have approved 94 percent of bond issues that the state has proposed, compared to 50 percent approval in the nine years before the Commission was established. Second, when Kentucky recently decided to revamp its capital planning process, it examined the processes used in many other states and eventually selected the New Jersey model.

Appendix A

Descriptions of Federal Capital Grant Programs

Function	Agency	Program	Federal Matching Rates and Requirements
Transportation			
Highways			
Prior to 1992	Federal Highway Administration	Interstate	90 percent of new construction costs for roads on the Interstate network.
		Interstate 4R	90 percent of costs for resurfacing, restoring, rehabilitating, or reconstructing the Interstate System. If needs of Interstate roads are fully met, the money may be used for primary roads.
		Primary	75 percent of costs for non-Interstate major roads that serve as intrastate, regional, or cross-state linkages (258,000 miles of road eligible for program).
		Secondary	75 percent of costs for major rural roads (400,000 miles of road eligible for program).
		Urban	75 percent of costs for roads primarily serving urban areas (148,000 miles of road eligible for program).
		Bridge Replacement and Rehabilitation	80 percent of costs to restore or replace bridges.
1992 and Beyond	Federal Highway Administration	National Highway System	75 percent of construction costs for all types of projects, except for repair or improvement of the Interstate system which will remain at 90 percent. This program will encompass the existing Interstate system and portions of the current Urban, Primary, and Secondary systems (150,000 miles).
		Urban/Rural Program	60 percent of construction costs. This program will include the rest of the current Urban, Primary, and Secondary systems (700,000 miles). Recipients may use grants either for highway or mass transit projects.
		Bridge Program	75 percent of costs to restore or rehabilitate bridges.
		Toll Projects	A new program will provide up to 35 percent of the costs of toll projects and will encourage private participation in these projects.
Mass Transit Prior to 1992	Urban Mass Transit Administration	Discretionary Capital Grants	Up to 75 percent of costs, except projects to improve accessibility for the elderly or handicapped, which receive a 95-percent match. A state may increase the priority of a project by supplying more than 25 percent of the funding.
		Nonurbanized Formula Grant, Urban Formula Grant	80 percent of costs, except projects to improve accessibility for the elderly or handicapped, which are eligible for a 95-percent match. Both programs also offer grants for up to 50 percent of operating costs (Note: Estimates of mass transit grants reported in tables exclude operating grants.)
		Capital Assistance for the Elderly and Handicapped	For private nonprofit organizations that provide transportation services for the elderly or handicapped. Organizations apply through the state and receive grants for up to 80 percent of project costs.
1992 and Beyond	Urban Mass Transit Administration	Discretionary Capital Grants	Up to 50 percent of project costs for significant new transit investment projects and 60 percent of costs for other capital projects.
		Nonurbanized Formula Grant, Urban Formula Grant	60 percent of project costs. Both programs will continue to offer grants to cover up to 50 percent of operating costs, although no operating assistance will be available for urban areas with a population of 1 million or more.
		Capital Assistance for the Elderly and Handicapped	60 percent of project costs.
Aviation	Federal Aviation Administration	Airport Improvement Program	For primary commercial service airports (those enplaning more than 0.25 percent of passengers nationwide), the matching rates are 75 percent for airport development, terminal development, master planning and noise compatibility planning, and 80 percent for noise compatibility program implementation. For all other public use airports, the rates are 90 percent for airport development, planning, noise compatibility planning and noise compatibility program implementation, and 75 percent for terminal development.
Wastewater Treatment	Environmental Protection Agency	Construction Grants	55 percent of costs for projects employing conventional technologies and 75 percent of costs for projects using innovative technologies. These grants will be phased out by 1991.
	Farmers Home Administration	Rural Water and Waste Disposal	Provides grants for rural waste disposal systems. These grants are distributed on a formula basis with no matching requirements specified.
Water Supply	Farmers Home Administration	Rural Water and Waste Disposal	Provides grants for rural water systems. These grants are distributed on a formula basis and have no matching requirements.
Community and Regional Development	Dept. of Housing and Urban Development	Community Development Block Grants, Urban Block Grants	Provides grants to promote development, which can be used for either capital or operating expenses. No matching requirements are specified, rather the grants are distributed by formula.

Source: U.S. Office of Management and Budget and U.S. General Services Administration. 1990. *Catalog of Federal Domestic Assistance 1990.* Washington, D.C.: Government Printing Office.

Appendix B: Tables not in numerical order because of space limitations.

Appendix Table B1
Federal Capital Grants to Massachusetts by Function, Fiscal Years 1980 and 1989
Millions of 1989 Dollars

Function	1980	1989
Total	$787.0	560.6
Transportation	466.0	351.8
Highway	220.8	234.1
Mass Transit	231.2	107.9
Aviation	14.0	9.8
Wastewater Treatment	135.3	92.4
Water Supply	2.1	1.2
Other	183.5	115.2
Addendum:		
Grants as a Percent of Capital Outlays		
Massachusetts	46.1	20.0
U.S. Total	35.7	21.8

Sources: U.S. Department of the Treasury, 1981, *Federal Aid to States, Fiscal Year 1980;* U.S. Bureau of the Census, 1990, *Federal Expenditures by State for Fiscal Year 1989*, Table 2; U.S. Office of Management and Budget, *Budget of the United States Government, Fiscal Year 1991* and *Fiscal Year 1989, Historical Tables*, Tables 9.5 and 12.3.

Appendix Table B5
Infrastructure Quality, Selected Years

	Percentage Rated Deficient			
	Highway Mileage		Bridges	
States	1982	1989	1980	1988
United States Average	13.7	9.5	40.5	41.3
Massachusetts	3.2	9.8	18.8	38.7
Other New England States	11.9	12.5	37.4	46.8
Connecticut	9.6	3.6	33.2	64.0
Maine	11.7	13.5	18.0	29.7
New Hampshire	13.5	25.2	51.3	43.7
Rhode Island	32.4	27.8	17.0	19.4
Vermont	6.9	5.6	54.5	49.2
High Technology States	13.2	7.1	36.1	33.6
Arizona	7.0	20.0	5.1	7.3
California	9.7	10.3	23.4	25.7
Maryland	8.4	5.4	17.7	40.9
North Carolina	7.8	5.4	75.7	52.5
Texas	19.6	5.0	36.7	34.2
Washington	8.5	2.2	12.0	27.0
Industrial States	16.2	8.3	38.0	40.9
Illinois	9.7	3.2	37.0	28.9
Michigan	21.5	9.9	36.0	31.3
New Jersey	.4	11.4	25.0	35.1
New York	11.3	2.5	58.4	68.2
Pennsylvania	26.4	17.0	24.8	39.7

Source: U.S. Department of Transportation, *Highway Statistics 1989*, and *Highway Statistics 1982*, Table HM-63; U.S. Department of Transportation, 1989, *The Status of the Nation's Highways and Bridges: Conditions and Performance*, Tables 4A and 4B; U.S. Department of Transportation, 1981, *Highway Bridge Replacement and Rehabilitation Program, Second Annual Report to Congress*, Tables 5A and 5B.

Appendix Table B4
Public and Private Capital Stocks Per Capita
1989 Dollars

	Public Capital Stock per Capita			Private Capital Stock per Capita		
	1970	1980	1989	1970	1980	1989
United States Average	5,940	6,726	6,860	12,501	16,026	18,804
Massachusetts	4,652	6,323	6,598	8,045	10,931	16,510
Other New England States	5,828	6,432	5,907	9,564	11,937	15,609
Connecticut	6,313	7,153	6,672	9,582	12,389	17,292
Maine	4,702	5,366	5,094	11,308	13,333	15,133
New Hampshire	5,027	5,710	4,961	9,709	11,242	14,609
Rhode Island	5,480	5,634	5,249	6,913	9,136	11,418
Vermont	7,106	7,175	6,293	10,964	12,560	16,353
High Technology States	6,755	6,942	6,706	13,827	16,770	19,162
Arizona	6,848	7,272	8,057	15,915	15,024	16,639
California	7,774	7,141	6,046	10,450	13,043	16,669
Maryland	5,880	8,052	8,002	8,943	11,583	14,469
North Carolina	4,000	4,968	5,077	10,331	13,494	16,979
Texas	5,763	6,068	6,715	22,958	26,219	26,313
Washington	9,121	10,273	10,666	13,249	16,681	18,379
Industrial States	5,888	7,286	7,216	10,550	14,259	17,010
Illinois	5,667	6,773	6,871	12,471	17,286	18,603
Michigan	6,070	6,779	6,489	11,514	15,588	17,945
New Jersey	4,178	5,521	5,855	9,747	12,424	17,358
New York	7,079	9,240	9,207	9,050	12,224	15,830
Pennsylvania	5,160	6,382	6,017	10,820	14,460	16,285

Source: Authors' calculations, see Munnell (1990b).

Appendix Table B2

State and Local Capital Spending Per Capita, Fiscal Years 1970 to 1989

1989 Dollars

State	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989
United States Average	483	490	471	453	435	430	422	385	348	360	367	365	346	348	350	373	407	433	440	450
Massachusetts	436	452	485	489	443	353	348	273	269	356	298	298	297	323	312	349	377	401	448	474
Other New England States	479	510	423	412	390	341	323	293	259	255	251	228	217	228	233	275	296	351	392	488
Connecticut	527	583	478	468	459	378	306	239	231	248	258	238	212	227	233	292	331	418	428	611
Maine	356	405	340	353	351	288	305	386	316	280	242	212	217	232	231	227	261	307	372	396
New Hampshire	453	485	440	421	357	417	485	440	351	257	263	238	228	243	225	252	247	336	374	419
Rhode Island	409	328	275	252	202	220	270	253	193	207	218	194	205	192	215	274	263	231	355	366
Vermont	621	670	525	485	455	321	305	253	271	324	264	251	246	265	294	327	318	305	326	340
High Technology States	518	499	472	449	438	435	439	402	400	413	418	402	386	383	369	395	452	490	490	470
Arizona	554	568	644	581	694	632	602	639	628	686	626	574	523	622	575	641	771	810	806	793
California	558	507	461	412	412	393	376	316	315	275	295	304	281	293	295	331	369	395	407	407
Maryland	499	576	536	548	599	601	574	518	484	440	512	411	442	409	425	395	481	506	543	532
North Carolina	354	349	333	323	312	338	404	338	311	379	289	253	225	221	226	301	353	402	417	398
Texas	457	430	455	443	414	416	408	385	392	458	485	451	423	427	431	433	540	578	540	500
Washington	732	785	634	691	525	583	717	791	822	907	838	881	934	781	582	595	499	612	625	543
Industrial States	482	514	506	488	460	423	391	345	279	297	304	320	298	305	310	326	365	398	429	443
Illinois	416	448	445	447	376	373	405	385	301	310	341	371	283	288	305	331	356	383	388	361
Michigan	438	400	432	398	387	392	379	324	272	298	331	278	232	243	258	232	264	271	372	336
New Jersey	437	441	400	380	389	307	296	281	226	274	256	311	283	269	268	271	386	427	445	466
New York	558	640	670	664	630	560	437	395	338	326	322	357	394	393	413	455	486	549	569	624
Pennsylvania	489	511	431	392	382	358	373	286	310	252	249	257	232	260	225	234	258	268	292	323

Source: U.S. Bureau of the Census, *Government Finances*, various years.

Appendix Table B3

State and Local Capital Spending as a Percent of Gross State Product, Fiscal Years 1970 to 1989

1989 Dollars

State	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989
United States Average	3.1	3.3	3.2	2.9	2.8	3.1	3.0	2.6	2.3	2.4	2.6	2.5	2.2	2.2	2.1	2.2	2.3	2.3	2.3	2.3
Massachusetts	2.8	3.0	3.2	3.2	2.9	2.7	2.6	1.9	1.9	2.5	2.2	2.1	1.9	2.0	1.8	1.9	1.9	1.9	2.1	2.1
Other New England States	3.1	3.5	2.9	2.8	2.7	2.6	2.5	2.1	1.8	1.8	1.9	1.7	1.5	1.5	1.4	1.5	1.6	1.8	1.9	2.3
Connecticut	3.0	3.5	2.9	2.8	2.8	2.5	2.0	1.5	1.4	1.6	1.7	1.5	1.2	1.3	1.2	1.4	1.5	1.8	1.8	2.4
Maine	2.9	3.5	2.9	3.0	3.0	2.7	2.9	3.3	2.7	2.4	2.2	1.9	1.8	1.8	1.7	1.6	1.8	2.0	2.3	2.3
New Hampshire	3.5	3.9	3.6	3.4	2.9	3.7	4.3	3.7	2.9	2.1	2.2	1.9	1.7	1.7	1.4	1.5	1.4	1.8	1.9	2.0
Rhode Island	2.9	2.4	2.0	1.8	1.5	1.8	2.2	2.0	1.5	1.7	1.8	1.6	1.6	1.4	1.5	1.8	1.7	1.4	2.1	2.1
Vermont	4.7	5.2	4.1	3.7	3.5	2.9	2.7	2.1	2.3	2.7	2.3	2.1	1.9	2.0	2.0	2.2	2.0	1.8	1.9	1.9
High Technology States	3.3	3.3	3.1	2.8	2.8	3.0	2.9	2.5	2.5	2.6	2.7	2.6	2.3	2.2	2.1	2.1	2.4	2.5	2.5	2.3
Arizona	4.1	4.3	4.8	4.1	4.9	4.9	4.8	4.8	4.6	5.0	4.8	4.4	3.8	4.6	4.0	4.2	4.8	4.8	4.7	4.6
California	3.2	3.0	2.8	2.4	2.4	2.6	2.3	1.9	1.8	1.6	1.8	1.8	1.6	1.7	1.6	1.7	1.8	1.9	1.9	1.9
Maryland	3.5	4.1	3.8	3.8	4.2	4.6	4.3	3.7	3.4	3.2	3.9	3.1	3.1	2.8	2.7	2.4	2.8	2.8	2.9	2.8
North Carolina	2.6	2.7	2.5	2.3	2.2	2.7	3.2	2.6	2.3	2.9	2.4	2.1	1.7	1.6	1.5	1.9	2.2	2.4	2.4	2.2
Texas	3.1	3.0	3.1	3.0	2.7	2.9	2.6	2.3	2.3	2.7	3.0	2.6	2.2	2.2	2.2	2.2	2.7	3.0	2.7	2.5
Washington	4.6	5.3	4.3	4.5	3.4	4.1	4.8	5.0	5.2	5.7	5.5	5.9	6.0	4.9	3.5	3.5	2.9	3.3	3.4	2.9
Industrial States	2.8	3.1	3.0	2.9	2.8	2.9	2.6	2.2	1.7	1.9	2.1	2.2	1.9	1.9	1.8	1.8	2.0	2.0	2.2	2.2
Illinois	2.3	2.6	2.6	2.5	2.1	2.3	2.5	2.2	1.7	1.8	2.1	2.4	1.7	1.7	1.8	1.8	1.9	2.0	2.0	1.8
Michigan	2.4	2.5	2.6	2.4	2.3	2.8	2.7	2.1	1.7	1.9	2.2	2.0	1.6	1.7	1.7	1.4	1.5	1.5	2.1	1.8
New Jersey	2.6	2.7	2.5	2.3	2.4	2.1	2.0	1.8	1.5	1.8	1.8	2.1	1.7	1.6	1.5	1.4	1.9	2.0	2.0	2.0
New York	3.0	3.6	3.8	3.7	3.6	3.7	2.8	2.5	2.1	2.1	2.1	2.4	2.4	2.3	2.2	2.3	2.4	2.5	2.6	2.7
Pennsylvania	3.2	3.5	3.0	2.7	2.6	2.7	2.8	2.0	1.5	1.8	1.9	2.0	1.6	1.8	1.5	1.5	1.6	1.6	1.7	1.9

Source: U.S. Bureau of the Census, *Government Finances*, various years; U.S. Bureau of Economic Analysis, Gross State Product, Machine Readable Data.

Appendix Table B6

Capital Spending of the State and the Four Major Authorities, Fiscal Years 1980 to 1999

Millions of 1989 Dollars

	State				Authorities											
					MWRA											
	Central Artery		Other		Harbor Cleanup		Other		Massport		MBTA		Turnpike		Total	
Year	Federal	State	Federal	State	Federal	State/ Local	Federal	State/ Local	Federal	State/ Local	Federal	State/ Local	Federal	State/ Local	Federal	State/ Local
1980	0	0	226	283	0	0	0	0	9	32	222	176	0	13	457	504
1981	0	0	252	373	0	0	0	0	9	55	236	186	0	11	496	625
1982	0	0	219	407	0	0	0	0	7	42	223	176	0	16	449	642
1983	0	0	225	483	0	0	0	0	8	41	224	177	0	15	458	716
1984	0	0	155	394	0	0	0	0	4	30	222	176	0	19	381	619
1985	0	0	188	461	0	0	0	0	5	56	218	172	0	32	411	721
1986	0	0	228	458	0	0	11	14	4	76	176	139	0	32	420	720
1987	0	0	249	581	0	0	14	33	16	46	167	132	0	32	447	823
1988	0	0	270	656	0	0	46	78	5	43	158	196	0	37	480	1010
1989	0	0	291	938	0	0	67	69	7	46	120	226	0	37	485	1317
1980–89	0	0	2305	5034	0	0	138	193	75	466	1965	1757	0	245	4483	7696
1990	28	3	291	925	18	52	38	94	10	68	95	284	0	69	480	1495
1991	273	30	291	925	15	133	11	151	10	68	130	220	0	100	728	1627
1992	409	45	291	925	19	347	19	162	10	68	142	188	0	95	890	1831
1993	481	76	291	925	15	416	38	155	10	231	150	180	0	80	985	2062
1994	615	97	291	925	9	469	15	122	10	263	160	175	0	66	1100	2117
1995	759	119	291	925	9	375	4	193	10	290	170	190	0	57	1243	2149
1996	727	114	291	925	10	195	0	207	10	357	180	205	0	55	1217	2058
1997	572	90	291	925	6	113	1	292	10	327	180	205	0	49	1058	2002
1998	371	58	291	925	7	146	1	258	10	256	180	205	0	48	860	1897
1999	90	14	291	925	7	133	1	271	10	155	180	205	0	31	578	1735
1990–99	4325	648	2905	9250	115	2378	128	1905	100	2083	1567	2057	0	650	9139	18971

Note: MWRA spending figures from 1986 to 1989 include $110 million of expenditures on the Harbor Cleanup project.

Source: Authority annual reports; Authority Budget/Planning Offices; Central Artery/Third Harbor Tunnel Public Information Office; Commonwealth of Massachusetts, *Annual Financial Report;* Background data supplied to Crozier Commission.

The authors would like to thank Edith B. Page of the Office of Technology Assessment, and Alan A. Altshuler, Director of the Taubman Center for State and Local Government, The John F. Kennedy School of Government, for valuable comments. Thanks also go to Donald E. Muterspaugh and Donna Hirsch of the Bureau of the Census, Kathy Emrich of the Commonwealth of Massachusetts Comptroller's Office and Ranch Kimball of Boston Consulting Group for their help in providing data, and to the Congressional Budget Office for the use of their Infrastructure Database.

[1] The federal government's direct expenditure on nonmilitary public capital in 1989 amounted to $19 billion.

[2] The following discussion focuses on federal grant programs, but the federal government also provides limited aid to states for physical capital investment in the form of loans. However, the only area of capital investment where loans represent a significant portion of total aid is water supply. Data compiled by the Congressional Budget Office show that in 1989 federal loans to states for the construction of water supply facilities represented 76 percent ($262 million out of $343 million) of all the aid in this area. These loans originate with the Farmers Home Administration and are targeted at rural communities. Since water supply is a very small portion of federal aid for capital investment, ignoring the loan programs does not significantly alter the relative importance of the federal government's role in financing capital investment.

[3] The actual assessments are made by the states based on EPA standards, and the EPA generally accepts the states' estimates. Until 1985, EPA provided 75 percent of the capital costs for systems employing conventional technologies and 85 percent for those based on innovative technologies. Since 1985, the matching rates have been lowered to 55 percent and 75 percent, respectively.

[4] Generally, Congress ends up appropriating somewhat less than the authorized amount.

[5] In this article (and throughout the report of which it is a part), Massachusetts is compared with a group of 16 similar states. The group includes the other New England states (Connecticut, Maine, New Hampshire, Rhode Island, and Vermont); six high technology states (Arizona, California, Maryland, North Carolina, Texas, and Washington); and five mature industrial states (Illinois, New Jersey, New York, Michigan, and Pennsylvania).

[6] This may seem surprising since Massachusetts began the period with the lowest level of per capita public capital, and its annual per capita investment over the period 1970 to 1989 was well below all but that for the other New England states. The explanation lies in the enormous difference in the rate of population growth between Massachusetts and other states. For example, the stock of public capital in the high technology states increased roughly 43 percent over the 1970–89 period, but this was inadequate to keep pace with the 44 percent increase in population; hence real per capita public capital declined by 1 percent. In Massachusetts, by contrast, the stock of public capital increased 46 percent, while population grew only 4 percent; as a result, per capita public capital rose 42 percent.

[7] Two basic sources provide data on capital spending by the state government: *Government Finances* published by the Census Bureau and Massachusetts' *Annual Financial Report*. The figures reported in these two sources are quite different. The 1989 Census Bureau figure for state government capital spending in Massachusetts is $1,093 million. Massachusetts' 1989 *Annual Financial Report* records capital expenditures of $1,230 million.

At first glance, these figures seem significantly different but not wildly divergent. A significant problem arises, however, because the Census Bureau includes as state government entities several public authorities whose expenditures do not pass through the state budget process, and thus do not appear in state documents. The Census Bureau performed a tabulation of these off-budget entities for Massachusetts. In 1989, these entities spent $395 million on capital outlays which, when subtracted from the Census Bureau figure for total state government capital spending, yields a value of $698 million in on-budget capital spending for the state government. This figure *is* wildly divergent from the $1,230 million recorded by the state.

After receiving some detailed state reports and having several conversations with the Census Bureau, a reconciliation of these figures was constructed. The major differences are discussed here, with a somewhat more detailed table appearing below.

First, a large portion of Massachusetts' state spending ($414 million) is classified as state aid/intergovernmental expenditure by the Census Bureau and is recorded as local, rather than state, government expenditure since localities are the final disbursing units. Second, some spending considered capital spending by the state is classified as operating expenditure by the Census Bureau ($67 million). Third, the Census Bureau classifies some expenditures ($20 million) as transfers or other expenditures. Subtracting these Census Bureau reclassifications from the state figure yields a capital spending figure of $728 million, just 4 percent larger than the Census Bureau figure of $698 million.

One may be tempted to ask: Which figure is correct, the Census Bureau estimate of $698 million or the state's figure of $1,230 million? The answer is that both have their merits. The Census Bureau figure is the most appropriate one to use when looking at spending of all levels of government within the state since it displays spending by final disbursing unit, thus avoiding the problem of double counting that would occur if intergovernmental expenditures of the state were counted as both state and local expenditures. On the other hand, from the perspective of the state, the non-federal portion of the spending recorded on its audited books ($938 million) is the amount that matters. This is the amount of money it had to raise through bond issuance, and it is the debt service on this figure which appears in the operating budget.

Reconciliation of Census Bureau and Commonwealth of Massachusetts Capital Spending Data, Fiscal Year 1989, Thousands of Dollars

Census Bureau	
State government capital spending	$1,092,748
Off-budget entities	(394,655)
	698,093

Commonwealth of Massachusetts	Fund Type				
	General	Highway	Federal	Other	Total
Annual Report Capital Spending	680,234	118,970	291,409	139,206	1,229,819
Census classifies as state aid/ intergovern-mental expenditure	(322,409)	(16,034)	(3,108)	(72,889)	(414,440)
Census classifies as operating expenditure	(53,869)	(622)	0	(12,592)	(67,083)
Census classifies as transfer/ other expenditure	(17,401)	0	0	(2,688)	(20,089)
Total	286,555	102,314	288,301	51,037	728,207
Difference from Census Bureau Total					30,114
Difference as Percent of Census Bureau Total					4.3

[8] According to Section 17 of Chapter 581 of the Acts of 1980, the MBTA is supposed to have been submitting a prospective budget coinciding with the state's fiscal year since July 1, 1983. While the MBTA now keeps its books on a fiscal year corresponding to the Commonwealth's, it continues to calculate its net cost of service on a calendar-year basis to be appropriated retrospectively.

[9] The MBTA is descended from Massachusetts' first public authority, the Boston Transit Commission, which was created in 1894 and built the subway tunnels. In 1918, the Boston Transit Department replaced the Commission and inherited the unpaid tunnel bills. The Metropolitan Transit Authority replaced the Boston Transit Department in 1949 and inherited the tunnel bills and other unpaid bills. These were all transferred to the MBTA when it was created in 1964.

[10] In 1989 only 2.5 percent of Massachusetts' interstate highway mileage, which includes the Mass Turnpike, was rated deficient, compared to 9.3 percent nationwide. At Logan, only 20 of every 1000 takeoffs and landings were delayed in 1989, compared to an average of 31 for the nation's 22 major airports.

[11] Some movement appears already to have been made in this direction. *The Boston Globe* (10/13/90) reported that $4.5 million of operating expenses for the existing central artery was transferred from the Department of Public Works to the Turnpike Authority, after the Authority's counsel offered help during the latest round of budget cuts.

[12] Most documents cite the cost for the Central Artery project as $4.4 billion, which is measured in constant 1987 dollars. The Central Artery Project's Public Information Office provided annual cost estimates in 1989 dollars. The figure commonly cited as the price tag for the Harbor Cleanup project is $6.1 billion. This is the cost inflated to 1999 dollars. In constant 1990 dollars the cost is estimated at $2.8 billion, according to the MWRA. Converting this to 1989 dollars results in a cost estimate of $2.7 billion.

[13] The Lazard Frères approach and the methodology used in this study differ, and these differences are summarized below. First, the annual expenditures used in the Lazard Frères report represent incurred obligations, rather than cash expenditures, which produce larger costs for the early years of the project in the Lazard calculations.

Second, Lazard Frères assume somewhat different federal matching rates. For the approved portion of the project, they assumed a 90-percent federal match on 97 percent of project costs, while this study assumed a 90-percent match on all costs. For the unapproved 20 percent of the project, the Lazard assumption was an 85-percent match of 97 percent of the costs up to and including 1992, with the match dropping to 82.5 percent on 97 percent of total costs after 1992. This study assumed no expenditures on the unapproved portion prior to 1992, and a federal match of only 75 percent when expenditures began after 1992. The following table shows costs by project section, broken down into federal and state shares before and after 1992.

Reconciliation of Federal Funding Estimates for the Central Artery Project
Millions of 1989 Dollars

	FY1990–FY1999		
	Total	Federal	State
Lazard Frères:			
Unapproved	958	771	187
Approved	4,014	3,504	510
Total	4,972	4,275	697
Munnell & Cook:			
Unapproved	994	745	249
Approved	3,978	3,580	398
Total	4,972	4,325	648

	FY1990–FY1992			FY1993–FY1999		
	Total	Federal	State	Total	Federal	State
Lazard Frères:						
Unapproved	178	147	31	780	624	156
Approved	1,605	1,401	204	2,409	2,103	306
Total	1,783	1,548	235	3,189	2,727	462
Munnell & Cook:						
Unapproved	0	0	0	994	745	249
Approved	789	710	79	3,189	2,870	319
Total	789	710	79	4,183	3,615	568

Source: Lazard Frères and Co., 1990, *Financing Plan for the Central Artery/Third Harbor Tunnel*, Chapter 2.

[14] The initial problems leading to this federal court order developed because the treatment plants and the sewer system are very old and have been consistently operating beyond their capacity. In the early 1900s, the system was recognized as one of the best in the country, though all it did was collect wastewater which was then released into the harbor. The Nut Island primary treatment plant was completed in 1952, providing wastewater treatment for the first time, and by 1968 the Deer Island primary treatment plant was also operational.

Over the next decade, growth among the communities served, combined with neglect of maintenance due to inadequate funding, resulted in treatment far below federal requirements. The two current primary treatment plants remove only 50 percent of bacterial pollution from wastewater, while the federal government requires that 90 percent be removed.

By the early 1980s, the situation had substantially deteriorated. Boston's treatment plants on Deer and Nut Islands experienced several instances of pump problems and capacity constraints, which resulted in the dumping of billions of gallons of untreated sewage into the harbor. Several lawsuits were filed against the Metropolitan District Commission (MDC), which culminated in the July 1985 finding that the MDC and the Massachusetts Water Resources Authority (MWRA) as its successor agency were liable for violating federal water pollution laws and responsible for remedying the intolerable situation. Thus years of neglect, substantial undercapacity, and outdated treatment technologies have created one of the most polluted waterways in the nation.

[15] An analysis by Cambridge Systematics indicates that, in its peak year, the Central Artery/Tunnel project will require 4,600 construction workers and will create another 10,500 jobs indirectly through the multiplier effects. A similar study performed by Cape Ann Economics for Associated Industries of Massachusetts produced more conservative estimates of 7,000 total jobs in the average year and 10,000 total jobs in the peak year. Cambridge Systematics also performed a similar analysis of the Harbor Cleanup project which estimated that, in its peak year, the project will generate 3,600 construction jobs and 6,200 other jobs for a total of 9,800 jobs.

References

Advisory Commission on Intergovernmental Relations. 1987. *Significant Features of Fiscal Federalism, 1988 Edition*, vol. 1 (December). Washington, D.C.: ACIR.

Aschauer, David A. 1989. "Is Public Expenditure Productive?" *Journal of Monetary Economics*, vol. 23, no. 2, pp. 177–200.

Boston Consulting Group. 1990. "Massachusetts Public Authorities: Information and Background Data." Background Report to Governor's Management Task Force. Boston: Boston Consulting Group.

Cambridge Systematics, Inc. 1989. "Economic Benefits of the Central Artery/Third Harbor Tunnel." Cambridge: Cambridge Systematics, Inc.

———. 1990. "Economic Impact of the Boston Harbor Project." Cambridge: Cambridge Systematics, Inc.

Caro, Robert A. 1974. *The Power Broker: Robert Moses and the Fall of New York*. New York: Alfred A. Knopf, Inc.

Central Artery/Tunnel Project, Public Information Office. 1990. Information Packet.

Commonwealth of Massachusetts. 1989. *The Governor's Budget Submission for Fiscal Year 1990, Volume II, Capital Outlay.*

———. 1990a. *Bond Prospectus*, March 1, June 1.

———. 1990b. *The Governor's Budget Submission, Fiscal Year 1991, Volume II, Capital Outlay.*

Commonwealth of Massachusetts, Executive Office for Adminis-

tration and Finance, Division of Capital Planning and Operations. 1985. *Capital Budgeting Procedures Manual FY 1987.*
Commonwealth of Massachusetts, Executive Office for Administration and Finance, Office of the Comptroller. Various years, 1980 to 1989. *Annual Financial Report.*
Commonwealth of Massachusetts, Senate Committee on Ways and Means. 1985. "State Authorities: The Fourth Branch of Government?"
———. 1989. "State Authorities: Public Funds for Public Purposes."
Congressional Budget Office. 1988. *New Directions for the Nation's Public Works* (September). Washington, D.C.: Government Printing Office.
———. 1990. Infrastructure Database, Machine Readable Data.
Executive Office of the President, Office of the Press Secretary, 1991. "Fact Sheet: The Surface Transportation Assistance Act of 1991." (February 13.)
Federal Aviation Administration, Air Traffic Operations Service. 1989. "Air Traffic Activity and Delays Report for December 1989" (December). Washington, D.C.: F.A.A.
Governor's Management Task Force. 1990. *Massachusetts: Managing Our Future* (Crozier Commission Report).
Gramlich, Edward M. 1990. "Financing Infrastructure Investment: Should Money Be Thrown at the Third Deficit?" Presented at the Annual Economic Conference of the Federal Reserve Bank of Boston, June 28–29.
Hornbeck, John F. 1988. "The Airport Improvement Program: Selected Economic and Legislative Issues." Library of Congress, Congressional Research Service, Report for Congress (October). Washington, D.C.: Library of Congress.
———. 1989. "Federal Policy and Financing Alternatives for Surface Transportation Infrastructure." Library of Congress, Congressional Research Service, Report for Congress (November). Washington, D.C.: Library of Congress.
———. 1990a. "Maintaining Highway and Bridge Investments: What Role for Federal Grant Programs?" Library of Congress, Congressional Research Service, Report for Congress, (May). Washington, D.C.: Library of Congress.
———. 1990b. "Transportation Trust Funds: Economic and Policy Issues." Library of Congress, Congressional Research Service, Issue Brief (July). Washington, D.C.: Library of Congress.
Lazard Frères and Co. 1990. *Financing Plan for the Central Artery/ Third Harbor Tunnel: Report to the Executive Office of Transportation and Construction* (November).
Ledebur, Larry C., William Hamilton, Deborah Matz and Lori Anderson. 1986. "Federal and State Roles in Infrastructure." Background Report to the National Council on Public Works Improvement. Arlington, VA: Aslan Institute.
Malone, M. E. and Jack Sullivan. 1990. "Dukakis Details Damage Done by Latest Round of Budget Cuts." *The Boston Globe,* October 13, p. 1.
Massachusetts Bay Transportation Authority. 1989. *Annual Financial Report.* Boston: Massachusetts Bay Transportation Authority.
Massachusetts Port Authority. Various years, 1980 to 1989. *Annual Report.* Boston: The Authority.
Massachusetts Special Commission on Infrastructure Finance. 1989. *A Survey of Massachusetts Infrastructure Needs for the 1990s.* Boston: The Commission.
Massachusetts Turnpike Authority. 1989. *Annual Report.* Boston: The Authority.
Massachusetts Water Resources Authority. Various years, 1986 to 1989. *Annual Report.* Boston: The Authority.
———. 1990. Information Packet on Harbor Cleanup Project.
Munnell, Alicia H. 1990a. "Why Has Productivity Growth Declined? Productivity and Public Investment." Federal Reserve Bank of Boston, *New England Economic Review,* (January/February), pp. 3–22.
Munnell, Alicia H. with the assistance of Leah M. Cook. 1990b. "How Does Public Infrastructure Affect Regional Economic Performance?" Presented at the Annual Economic Conference of the Federal Reserve Bank of Boston, June 28–29.
National Conference of State Legislatures. 1987. *Capital Budgeting and Finance: The Legislative Role.* Denver: National Conference of State Legislatures.
New Jersey Commission on Capital Budgeting and Planning. 1990. *State of New Jersey, Annual Capital Improvement Plan.* Trenton, NJ: The Commission.
Porter, Douglas R., Ben C. Lin and Richard B. Peiser. 1987. "Special Districts: A Useful Technique for Financing Infrastructure." Washington, D.C.: The Urban Land Institute.
U.S. Bureau of the Census. 1990. *Federal Expenditures by State for Fiscal Year 1989.* Washington, D.C.: Government Printing Office.
———. Various years, 1969–70 to 1988–89. *Government Finances.* Washington, D.C.: Government Printing Office.
U.S. Bureau of Economic Analysis. 1990. Gross State Product, Machine Readable Data.
U.S. Department of Transportation. 1990. *Moving America: New Directions, New Opportunities* (February). Washington, D.C.: Government Printing Office.
U.S. Department of Transportation, Federal Highway Administration. 1981. *Highway Bridge Replacement and Rehabilitation Program, Second Annual Report to Congress.* Washington, D.C.: Government Printing Office.
———. 1983. *Highway Statistics 1982.* Washington, D.C.: Government Printing Office.
———. 1989a. *Highway Statistics 1989.* Washington, D.C.: Government Printing Office.
———. 1989b. *The Status of the Nation's Highways and Bridges: Conditions and Performance.* Washington, D.C.: Government Printing Office.
U.S. Department of Transportation, Urban Mass Transportation Administration. 1981. *National Urban Mass Transportation Statistics: First Annual Report, Section 15 Reporting System* (May). Washington, D.C.: Government Printing Office.
———. 1989. *National Urban Mass Transportation Statistics: 1988 Section 15 Annual Report* (December). Washington, D.C.: Government Printing Office.
———. 1991. "Surface Transportation Act 1991: Media Briefing Highlights" (February).
U.S. Department of the Treasury. 1981. *Federal Aid to States, Fiscal Year 1980.* Washington, D.C.: Government Printing Office.
U.S. Office of Management and Budget. 1990. *Budget of the United States Government, Fiscal Year 1991.* Washington, D.C.: Government Printing Office.
———. 1988. *Historical Tables, Budget of the United States Government, Fiscal Year 1989.* Washington, D.C.: Government Printing Office.
U.S. Office of Management and Budget and U.S. General Services Administration. 1990. *Catalog of Federal Domestic Assistance.* Washington, D.C.: Government Printing Office.
U.S. Office of Technology Assessment. 1990. *Rebuilding the Foundations: A Special Report on State and Local Public Works Financing and Management.* Washington, D.C.: Government Printing Office.
Wheeler, Porter K. 1986. "The Allocation of Governmental Responsibilities in Constructing, Maintaining, and Financing Public Works." Background Report to the National Council on Public Works Improvement.

Federal Reserve Bank of Boston
P.O. Box 2076
Boston, Massachusetts 02106-2076

Address Correction Requested

Bulk Rate
U.S. Postage
PAID
Richmond, VA
Permit No. 930