THE COMOVEMENT OF STOCK PRICES*

ROBERT S. PINDYCK AND JULIO J. ROTEMBERG

We test whether comovements of individual stock prices can be justified by economic fundamentals. This is a test of the present value model of security valuation with the constraint that changes in discount rates depend only on changes in macroeconomic variables. Then, stock prices of companies in unrelated lines of business should move together only in response to changes in current or expected future macroeconomic conditions. Using a latent variable model to capture unobserved expectations, we find excess comovement of returns. We show that this excess comovement can be explained in part by company size and degree of institutional ownership, suggesting market segmentation.

I. INTRODUCTION

A fundamental question in financial economics is whether stock prices are equal to the present discounted value of expected future earnings. But the present value model cannot be tested absent a theory of discount rates. The most widely known tests of the present value model consider the case in which the discount rate is constant. (Examples are the volatility tests that began with the work of Shiller [1981] and LeRoy and Porter [1981].) That version of the model can easily be rejected and is not very interesting as a theoretical proposition. Almost all general equilibrium models involve changes in discount rates. For example, in the representative agent model with a single consumption good, the discount rate will change when consumption changes. More generally, we expect discount rates to depend on a variety of macroeconomic variables.

All present value models in which discount rates depend *only* on macroeconomic variables have a common implication: the prices of different stocks can move together only in response to common movements in earnings, or in response to common effects of changes in macroeconomic variables. Hence, the stocks of companies whose earnings are uncorrelated (e.g., because they are in unrelated industries) should move together only in response to changes in current or expected future macroeconomic conditions.

*We thank John Driscoll, Christine Chilvers, Patricia Craig, and Rebecca Emerson for research assistance; Andrew Alford, Fischer Black, Zvi Bodie, Eugene Fama, John Heaton, Bruce Lehmann, Andrew Lo, David Modest, Robert Shiller, three anonymous referees, and especially Andrei Shleifer for helpful comments. We also acknowledge financial support from the MIT International Financial Services Research Center, the MIT Center for Energy Policy Research, and the National Science Foundation.

The Quarterly Journal of Economics, November 1993

In this paper we test and reject this hypothesis. This rejection casts doubt on a broad class of models. The class includes not only the standard representative agent model, but also models like Mankiw's [1986], in which macroeconomic variables have important distributional effects. Thus, many models with incomplete markets are being tested alongside standard complete market models.

Our study is based on earnings and returns data for groups of selected companies. The companies in each group are unrelated according to two criteria. First, they operate in different lines of business: neither do they produce similar goods or services; nor do they have important vertical relations with each other. Second, their earnings (normalized by nominal GNP) are not significantly correlated. Both of these criteria still allow for correlations among these companies' earnings, but only insofar as those correlations are mediated through correlations with macroeconomic variables. Therefore, the stock returns of these companies should be correlated only to the extent that they are correlated with variables related to economywide earnings or discount rates. We test whether this is indeed the case.

To conduct our tests, we first run OLS regressions of stock returns on current and lagged values of macroeconomic variables. These regressions are similar to those run by Burmeister and McElroy [1988]; Chen, Roll, and Ross [1986]; McElroy and Burmeister [1988]; and others who tested the Arbitrage Pricing Theory (APT) by focusing on observable factors. We then test whether the residuals of these regressions are correlated across firms. We first run our regressions excluding the return on the market as an independent variable, on the grounds that any correlations between individual returns and a market index should be due to the effects of economic variables. We then include either the lagged dividend-price ratio for the market—a variable that has been shown (e.g., by Fama and French [1988]) to have predictive power for stock returns—or the return on the market itself. We find that in all cases the regression residuals are highly correlated for every group of companies.

One problem with this approach is that agents' expectations of future macroeconomic variables may be based on more than the current and lagged values of those variables. Any additional information that agents have about the future of the economy should be reflected in the stock prices of many firms, and hence can lead to a spurious finding of excess comovement in the least squares regressions. Unfortunately, this problem cannot be eliminated by running regressions of current returns on *future* values of macroeconomic variables, as is done, for example, by Chen, Roll, and Ross [1986]. The reasons is that agents can base their investment decisions only on expectations of these variables. The realizations of these variables represent market expectations plus expectational errors, and expectational errors can introduce spurious comovement in regressions of current returns on future values of macroeconomic variables.

Following our earlier work on commodity prices [Pindyck and Rotemberg, 1990], we account for expectations through the use of latent variables, which represent unobserved forecasts of macroeconomic variables. Our model then becomes a MIMIC (multiple indicator multiple cause) model. The "indicators," i.e., the variables affected by the latent variables, include both the individual stock returns and the actual future values of the macroeconomic variables. The "causes" are any variables useful in forecasting macroeconomic variables. In this case, the causes are current and past values of some economic variables (e.g., the money supply and oil prices), as well as the stock market itself.[1]

We use two latent variables, the first representing forecasts of future real GNP growth, and the second representing forecasts of future inflation. Because nominal interest rates are explanatory variables in our equations, the latent variable for inflation captures the effects of real discount rates. These latent variables turn out to be significant explanators of stock returns. Nonetheless, for each group of companies the unexplained movements in returns remain excessively correlated. This is inconsistent with versions of the present value model in which discount rates depend only on macroeconomic variables.

These results suggest that security prices also depend on variables unrelated to aggregate conditions. One possibility is that they depend on variables that affect only particular groups of investors. Suppose that one such group receives news about its members' future income. This would affect their marginal utility of wealth. If financial markets are segmented so that certain securities are held predominantly by this group of investors, then this news will also affect the prices of these securities. In this example the present value model holds, but financial market segmentation

1. Fama [1981] and Fischer and Merton [1984] have provided evidence that the stock market is one of the best predictors of real economic variables.

can cause the value of a security to be different for different investors. Another possibility is that markets are segmented and that groups of investors change their beliefs about future dividends or stock prices for extraneous reasons. Such herding behavior would then affect the securities that are held disproportionately by these groups.

To see whether market segmentation of this sort can explain our results, we examine whether comovement is especially pronounced among stocks held by a particular segment, namely institutional investors. We find that the extent to which stocks are predominantly held by institutional investors indeed accounts for some of the comovement. Thus, market segmentation appears to have a role in explaining our results.

The early antecedents of this work are King [1966] and Meyers [1973]. King argues that most of the correlation in returns is attributable to industry effects.[1] However, he also finds significant correlation among the factors that represent industries. Using a principal components analysis, Meyers finds even less support for the idea that all return correlations are due to industry factors. (His principal components are not fully explained by industry effects.) More recently, Shiller [1989], studied comovements of returns between the U. S. and U. K. markets.[3] He showed that market averages as well as expected rates of return on market averages in these two countries move together. This finding constitutes strong evidence against the present value model with a constant discount rate. It is not inconsistent with the present value model, however, if one allows for plausible variations in discount rates; indeed, Shiller cannot reject the versions of his model where the discount rate varies with the rate of return on commercial paper.

Our work is also related to Huberman, Kandel, and Karolyi [1988], who show, in the context of Arbitrage Pricing Theory, that returns for firms of different size groups are affected by different factors. They suggest several possible explanations, one of which is that institutions tend to hold the stocks of larger companies. We show that there is indeed excess comovement for firms in different size groups, and we find strong support for the view that this is because institutional investors hold shares of large companies'

2. See also Lee and Vinso [1980], who study the correlation of returns of several oil companies.
3. See King and Wadhwani [1989] for an analysis of the interactions between markets in different nations at hourly frequencies.

stocks in their portfolios. Finally, Lee, Shleifer, and Thaler [1991] study the behavior of market price relative to net asset value for closed-end mutual funds. They show that the discounts on these funds move together, and also covary with the prices of small stocks. Their explanation is based on market segmentation; both closed-end funds and small stocks are held disproportionately by individual investors, who tend to hold similar sentiments.

The next section explains the underlying theory. We show that if stock prices represent present values of expected future earnings, and agents' utilities depend on a single consumption good, any systematic component of returns must be macroeconomic in nature. Section III explains our empirical methodology and the data that we use. Section IV shows that the earnings of the companies we have chosen are not highly correlated, but Section V shows that their returns are correlated even after accounting for the common effects of a variety of observable variables. Section VI presents test results based on our latent variable model. Section VII shows that the residual correlations from both our OLS and latent variable models can be partly explained by variables accounting for company size and degree of institutional ownership, suggesting that comovement may be due to common ownership of subsets of stocks. Section VIII concludes.

II. The Theory

Our test is based on the standard model in which the stock price of firm i at time t is the expected present discounted value of earnings:

$$P_{i,t} = E_t \sum_{j=0}^{\infty} \frac{A_{i,t+j}}{R_{t,t+j}}, \tag{1}$$

where E_t takes expectations conditional on information at time t, $P_{i,t}$ and $A_{i,t}$ represent, respectively, the share price and the earnings per share of company i at time t. $R_{t,t+j}$, is the rate at which investors discount back to t the earnings they receive at $t + j$ in a particular state of nature. We assume that, as in the Arrow-Debreu framework, there is only one rate at which investors discount earnings that accrue in a particular state.

Assuming for simplicity that earnings at t are paid out at t,

$$P_{i,t-1} = A_{i,t-1} + E_{t-1} \sum_{j=0}^{\infty} \frac{A_{i,t+j}}{R_{t-1,t+j}}.$$

Since, for any particular state of nature, the R's are known with certainty, $R_{t-1,t}R_{t,t+j} = R_{t-1,t+j}$ for $j > 1$. Combining these equations after dividing the latter one by $R_{t-1,t}$,

$$(2) \qquad \frac{P_{i,t}}{R_{t-1,t}} - (P_{i,t-1} - A_{i,t-1}) = \sum_{j=0}^{\infty} (E_t - E_{t-1}) \frac{A_{it+j}}{R_{t-1,t+j}}.$$

The expectation of the right-hand side of (2) at time $t - 1$ is zero. Now consider the return on an individual security, i.e., the ratio of the security's total payoff to its cost. Given our timing convention, this return for stock i from $t - 1$ to t, which we denote by $Q_{i,t}$, equals $P_{i,t}/(P_{i,t-1} - A_{i,t-1})$. Using (2),

$$(3) \qquad Q_{it} = R_{t-1,t} + R_{t-1,t} \frac{\Sigma_{j=0}^{\infty}(E_t - E_{t-1})A_{i,t+j}/R_{t-1,t+j}}{P_{i,t-1} - A_{i,t-1}}.$$

We now linearize equation (3) around the trajectories for expected future earnings and discount factors, $\{\overline{A}_{i,\tau}\}$ and $\{\overline{R}_{t-1,\tau}\}$. Because expectational revisions of $\overline{A}_{i,t+j}/\overline{R}_{t-1,t+j}$ are zero, we have

$$(4) \qquad \begin{aligned} Q_{i,t} \approx R_{t-1,t} + \left[\frac{\overline{R}_{t-1,t}}{\overline{P}_{i,t-1} - \overline{A}_{i,t-1}}\right] \sum_{j=0}^{\infty} \frac{\overline{A}_{i,t+j}}{\overline{R}_{t-1,t+j}} \\ \times \left[\frac{(E_t - E_{t-1})A_{i,t+j}}{\overline{A}_{i,t+j}} - \frac{(E_t - E_{t-1})R_{t-1,t+j}}{\overline{R}_{t-1,t+j}}\right]. \end{aligned}$$

Thus, a security's return can be approximated as a present value of expectational revisions of percentage deviations of earnings and discount rates from their mean paths.

We consider the class of models where the R's are related to macroeconomic variables. This class includes not only representative agent models but also models where macroeconomic shocks have important distributional consequences. In this class of models expectational revisions of future discount rates can depend only on macroeconomic variables, which we shall denote by M_t. Expectational revisions of future earnings will depend on macroeconomic variables, but can also depend on firm-specific variables that are uncorrelated with macroeconomic variables. We denote these latter variables by $z_{i,t}$. Equation (4) then implies that the return on security i can be written as

$$(5) \qquad Q_{i,t} \approx \overline{Q}_i + \sum_{j=0}^{\infty} [a_j E_t z_{i,t+j} + b_j E_t M_{t+j}].$$

One of our criteria for selecting groups of unrelated firms is that their earnings are uncorrelated; i.e., $E_t(z_{i,t}z_{j,t+s}) = 0$ for all t,s. Thus, the returns of any two companies, $Q_{i,t}$ and $Q_{j,t}$, can be correlated only through the common effect of the M_t's on earnings and discount factors. Also, the residuals of regressions of $Q_{i,t}$ and $Q_{j,t}$ on current and expected future M_t's should be uncorrelated. This is the basis for the tests that follow.

Note that the linearization of equation (4) is inconsistent with the CAPM in that the expected return on every security is equal to the discount rate $R_{t-t,t}$. We could have instead linearized around the expected path for $R_{t-1,t+j}$ and the expected path for the product $R_{t-1,t}A_{i,t+j}$. The return on security i would again be approximately equal to a sum of expectational revisions of earnings and expectational revisions of the discount rate, but the expected return would depend on the covariance of earnings with the discount rate, which in turn depends on the covariance of earnings with macroeconomic variables. Although consistent with the CAPM, that linearization is more complicated. In any case, it leads to the same conclusion as equation (4); correlations of returns can only be due to common correlations with macroeconomic variables.

III. Empirical Methodology and Data

Our tests are done in four steps. First, we form groups of companies in unrelated lines of business and test whether their earnings, normalized by nominal GNP, are indeed uncorrelated. Second, we run OLS regressions of the companies' returns on current and past values of macroeconomic variables, a market index, and the lagged dividend-price ratio, and then test whether the regression residuals are uncorrelated. Third, we estimate a latent variable factor model that accounts for unobserved expectations of future macroeconomic variables, and test whether the errors of this model are uncorrelated across companies, as the present value model would imply. Finally, we study the association between the correlation of returns residuals and both the size of the companies and the degree to which their shares are held by institutional investors.

Our choice of companies was constrained by the availability of quarterly data for earnings and returns. Quarterly earnings are from COMPUSTAT. We use the series for "Operating Income before Extraordinary Items" for two reasons. First, extraordinary items are typically unrelated to earnings from normal operations.

Second, they induce large outliers in the series for total earnings, which can dominate any measure of correlation constructed with total earnings. Unfortunately, COMPUSTAT contains complete histories since 1969 of "Operating Income before Extraordinary Items" for only a few companies. We included all of these companies in our sample, and added others for which this series was nearly complete, and for which it was possible to reasonably approximate the behavior of earnings by using the series for "Operating Income Inclusive of Extraordinary Items" to interpolate the missing data points. The resulting sample has 42 companies divided into six groups, labeled A to F, with seven companies in each. The companies and their principal lines of business are listed in Table I.

Quarterly returns data for these companies were obtained from CRSP by cumulating the three monthly returns corresponding to each quarter. Thus, the return in the first quarter is from the first working day after January 1 to the last working day on or before March 31. Data for the macroeconomic variables used in our tests are from CITIBASE. All of our tests use data from 1969:1 to 1987:4.

IV. The Behavior of Earnings

We begin by forming groups of companies whose main business activities are unrelated. The number of companies in each group cannot be too large, since otherwise there is bound to be overlap in their activities, which will show up as correlations in the z's. At the same time we want enough companies in each group to allow for sufficient degrees of freedom in our tests. A reasonable balance is achieved with groups of seven companies.

These groups are chosen to minimize the overlap and vertical connection between the companies' activities. Hence the components of earnings unrelated to macroeconomic factors should be uncorrelated, and we test whether this is the case. One way to do this is to regress earnings against macroeconomic variables and seasonal dummies, and then look at the correlation of the residuals. We did this, and found that the main macroeconomic variable which helps explain earnings is nominal GNP. Hence, we also computed correlations for the seasonally adjusted changes in normalized earnings—the ratio of earnings to nominal GNP. While both sets of correlations were small, we report those based on normalized earnings, which are somewhat smaller. (Allowing the elasticities of individual earnings with respect to GNP to differ

TABLE I
COMPOSITION OF GROUPS A, B, C, D, E, AND F

Symbol	Company name	Principal business
	Group A	
IP	International Paper Co.	paper products
CHV	Chevron	oil refining
NEM	Newmont Mining Corp	mining of nonferrous metals
SW	Stone and Webster	engineering architectural consulting
CQ	Communications Satellite	communications services
KR	Kroger	supermarkets
CHL	Chemical Bank	banking
	Group B	
AL	Alcan	aluminum smelting and refining
BS	Bethlehem Steel	steel works
CMB	Chase Manhattan	banking
MES	Melville Corp.	clothing stores
OM	Outboard Marine Corp	engines and turbines
STO	Stone Container	paper products
MUR	Murphy Oil Corp	oil refining
	Group C	
APD	Air Products and Chemicals	chemicals and industrial gases
DAL	Delta Air Lines	commercial air carrier
EFU	Eastern Gas and Fuel	gas company consortium
EBS	Edison Bros. Stores	shoe stores
MDP	Meredith Corp	television and print media
PD	Phelps Dodge Corp	mining of nonferrous metals
FI	First Interstate Bank	banking
	Group D	
BH	Belding Heminway	industrial threads and yarns
GG	Giant Group Ltd.	cement
KG	Kellogg Co.	food
PHI	Phillips Petroleum	oil refining
PIT	Pitney-Bowes	office machinery and supplies
TM	Times-Mirror	television and print media
ZEM	Zemex Corp.	tin, feldspar, iron powder mining
	Group E	
VUL	Vulcan Materials	construction materials
WEY	Weyerhauser	timber and forestry products
ING	Ingersoll-Rand	rock-drilling equipment
ZEN	Zenith	consumer electronics
MCD	McDonnell-Douglas	aircraft and defense contracting
DIS	Disney	entertainment products
UC	Union Carbide	industrial gases and chemicals
	Group F	
ARM	Armstrong World Industries	furniture
GNN	Great Northern Nekoosa	paper products
DV	Dover Corp	industrial equipment
TN	Tandy Corp	retail consumer electronics
WS	Weis Markets	supermarkets
NOR	Norwest Corp	banking
DU	DuPont	chemical and biomedical products

TABLE II
CORRELATIONS OF NORMALIZED EARNINGS CHANGES

Group A

	IP	CHV	NEM	SW	CQ	KR	CHL
IP	1.000						
CHV	0.007	1.000					
NEM	0.044	−0.105	1.000				
SW	−0.141	−0.125	0.028	1.000			
CQ	−0.156	0.070	−0.097	0.201	1.000		
KR	−0.284	0.099	−0.226	0.223	0.274	1.000	
CHL	0.104	0.071	−0.004	0.030	−0.164	−0.096	1.000

$\chi^s(21) = 30.08$

Group B

	AL	BS	CMB	MES	OM	STO	MUR
AL	1.000						
BS	−0.025	1.000					
CMB	0.077	0.137	1.000				
MES	0.036	−0.217	−0.092	1.000			
OM	−0.052	−0.164	0.064	0.206	1.000		
STO	0.251	0.149	−0.109	0.102	−0.016	1.000	
MUR	−0.103	0.009	0.004	−0.313	0.008	−0.178	1.000

$\chi^2(21) = 31.54$

Group C

	MDP	APD	PD	DAL	EFU	EBS	FI
MDP	1.000						
APD	0.021	1.000					
PD	−0.204	−0.144	1.000				
DAL	0.071	0.176	−0.048	1.000			
EFU	−0.224	−0.224	−0.121	−0.003	1.000		
EBS	−0.085	−0.090	0.076	−0.181	0.282	1.000	
FI	0.059	0.088	0.001	0.005	−0.016	−0.016	1.000

$\chi^2(21) = 29.68$

from one lowers the variance of unexplained earnings but, to the extent that the differences in these elasticity estimates are spurious, can raise their covariances.)

Normalized earnings are highly autocorrelated, making it difficult to compute the statistical significance of correlations in levels. Quarterly differences exhibit very little autocorrelation, so we calculate interfirm correlations for these first differences.

Table II shows the correlations of these first differences for each group of companies. Note that all of the correlations are low;

TABLE II
(CONTINUED)

	Group D						
	GG	PIT	ZEM	KG	BH	TM	PHI
GG	1.000						
PIT	0.033	1.000					
ZEM	0.126	0.076	1.000				
KG	−0.106	−0.231	0.014	1.000			
BH	0.192	0.053	0.201	0.035	1.000		
TM	−0.250	0.036	0.140	0.072	0.084	1.000	
PHI	−0.051	−0.038	0.047	0.018	0.012	0.020	1.000
			$\chi^2(21) = 21.16$				
	Group E						
	VUL	WEY	ING	ZEN	MCD	DIS	UC
VUL	1.000						
WEY	−0.051	1.000					
ING	0.033	0.048	1.000				
ZEN	−0.009	−0.162	−0.061	1.000			
MCD	−0.200	0.157	−0.057	−0.111	1.000		
DIS	0.002	−0.284	−0.076	0.281	−0.239	1.000	
UC	0.111	0.004	−0.017	−0.061	−0.177	0.066	1.000
			$\chi^2(21) = 26.18$				
	Group F						
	ARM	GNN	DV	TN	WS	NOR	DU
ARM	1.000						
GNN	0.004	1.000					
DV	−0.159	0.015	1.000				
TN	0.059	−0.209	0.248	1.000			
WS	−0.123	−0.137	0.003	0.133	1.000		
NOR	−0.010	−0.183	0.019	−0.031	0.285	1.000	
DU	0.228	−0.039	−0.047	−0.091	0.000	0.079	1.000
			$\chi^2(21) = 29.57$				

Note. Each entry is correl $(\Delta(E_i/Y), \Delta(E_\gamma/Y))$, where E_i/Y is earning of company i, Y is nominal GNP, and $\Delta(E_i/Y)$ is deseasonalized.

there are only two or three in each group that are above 0.22 in magnitude and hence individually significant at the 5 percent level.[4] We also test whether the 21 correlations for each set of companies are statistically significant as a group. This test is a likelihood ratio test of the hypothesis that the correlation matrix

4. With 75 quarterly observations, the critical ρ^* for significance at the 5 percent level satisfies $-75\log(1 - \rho^{*2}) = 3.841$ (where 3.841 is the critical value of $\chi^2(1)$), or $\rho^* = 0.223$.

for earnings is diagonal. As shown in Morrison [1967], the ratio of the restricted and unrestricted likelihood functions is $\lambda = |R|^{N/2}$, where $|R|$ is the determinant of the correlation matrix. The test statistic is therefore $-2 \log \lambda$, which is distributed as χ^2 with 21 degrees of freedom. Observe that the groupwise correlations are insignificantly different from zero for each set of companies; none of the reported χ^2 statistics are above the critical 5 percent level of 32.67.

One might argue that these low correlations are due in part to measurement error, even though we use operating income, which does not include depreciation, as our measure of earnings. To explore this possibility, suppose that the measured change in earnings of firm i at t, $\Delta A_{i,t}^m$, equals the change in true earnings, $\Delta A_{i,t}$ plus an error $\epsilon_{i,t}^a$. Suppose also that the variances of the ΔA's and of the ϵ^a's are the same for all companies and that the ϵ^a's are uncorrelated. Then, the correlation of true earnings changes equals the correlation of measured changes times $(1 + \sigma_\epsilon^2/\sigma_A^2)$, where σ_ϵ^2 is the variance of the measurement error while σ_A^2 is the variance of true earnings changes. How high would $\sigma_\epsilon^2/\sigma_A^2$ have to be to make our finding of insignificant correlation spurious? To answer this, we convert the χ^2 statistics in Table II into a correlation between two firms' earnings that would have the same level of significance as the respective statistic. This gives correlations of 0.19, 0.21, 0.19, 0.10, 0.15, and 0.19 for groups A through F. For such a correlation to be significant at the 5 percent level, it must equal 0.223; this would require the ratios $\sigma_\epsilon^2/\sigma_A^2$ to be 0.17, 0.06, 0.17, 1.23, 0.49 and 0.17 for groups A through F, respectively. These estimates assume that measurement error is uncorrelated across firms. It is more plausible that these errors are positively correlated, e.g., as firms distort reported earnings to take advantage of changes in tax laws. A positive correlation would bias our tests against the hypothesis of correlated earnings.

Because some individual correlations are significant, we also constructed a seventh group of companies (group G). These companies are also included in the other groups, but were chosen so that within this grouping no individual correlation exceeds 0.13. We report the results for this group at the end of Section VI.

Correlations of changes in normalized earnings may unduly emphasize high-frequency movements, and stock returns might covary because of low-frequency comovements of earnings. If normalized earnings are random walks, only the correlations of changes are relevant, but if earnings are stationary, the correlation of residuals in regressions explaining earnings with past earnings are more relevant. Thus, we also estimated a fourth-order autoregressive model for the normalized earnings of each company and calculated the correlation matrix for the residuals in each group. The χ^2 statistics for groupwise significance are 23.11, 37.01, 18.65, 19.26, 27.14, and 28.22 for groups A to F. Except for group B, none of these statistics are significant at the 5 percent level.

Low-frequency correlations could also result if the earnings of one company are correlated with the subsequent earnings of another company. To allow for this, we regressed the change in normalized earnings for each company on its own lagged earnings change, the lagged change in earnings of the other six companies in the group, as well as real GNP (current and lagged), a time trend, and the time trend squared. We then test whether, for each group, the lagged changes in the other six companies' earnings can be excluded from all seven regressions. The test statistics, distributed as $\chi^2(42)$ under the null hypothesis of no correlation, equal 17.43, 17.68, 20.19, 5.65, 12.90, and 18.06 for groups A to F, respectively. In this case the critical 5 percent level is 58.12, so that none of these statistics are significant. We conclude that normalized earnings for the companies in each group are uncorrelated, even allowing for a lag.

A final way to account for correlations at leads and lags is to examine annual and biannual changes. We thus constructed correlation matrices for fourth quarter to fourth quarter changes in normalized earnings, as well as eight quarter changes. These correlations are indeed higher. For the annual changes in earnings, the test statistics for groups A to F are 27.95, 25.26, 30.67, 29.97, 59.13, and 50.33. For the eight quarter changes they are 27.90, 64.26, 62.40, 32.19, 55.07, and 58.94. In the first case, two of the statistics are significant at the 5 percent level, while four are significant in the second case. This raises concern about the possibility of low-frequency earnings correlations. However, the paucity of data makes these statistics suspect. (There are only nine independent observations in the case of biannual correlations so that the seven by seven matrices of correlations are nearly singular.)

V. The Behavior of Returns

We next examine the correlations of returns. We calculated correlation matrices for the raw returns in each group, and χ^2 statistics to test groupwise significance. Returns are strongly positively correlated. The statistics are 153.52, 172.26, 173.07, 145.08, 224.90, and 262.42 for groups A to F, which are all highly significant. When we look at annual (fourth quarter to fourth quarter) returns, the correlations rise but the statistics fall to 93.28, 66.81, 81.40, 65.80, 71.23, and 130.12, which are still highly significant.

The correlation of raw returns is to be expected, since changing macroeconomic conditions can affect all returns by affecting expected future earnings. At issue is whether these returns remain correlated after controlling for macroeconomic effects. If investors' expectations of future macroeconomic conditions are based solely on current and past values of macro variables, simple OLS regressions can be used to filter out these effects. Then regressions of individual ex post returns on a sufficiently large set of current and lagged macro variables should lead to uncorrelated residuals.

We first run OLS regressions of returns on current and lagged values of five macroeconomic variables that could reasonably be expected to affect expected future earnings or discount rates: the log first difference of real GNP (Y), the log first difference of the GNP deflater (π), an index of the exchange value of the dollar against ten other currencies (EXCH), the ratio of aggregate corporate profits before tax (inclusive of depreciation) to nominal GNP (CPBT), and the three-month Treasury bill rate (TBILL). This allows both expected and unexpected changes in these variables to affect returns.

Macroeconomic variables explain very little of the variation in ex post returns. This can be seen from Table III, which shows test statistics for the hypothesis that a given macroeconomic variable (and its lag) can be excluded altogether from the seven regressions

TABLE III
χ^2 STATISTICS FOR GROUP EXCLUSIONS OF EXPLANATORY VARIABLES IN OLS REGRESSIONS

	Group					
Variable	A	B	C	D	E	F
Y	7.68	17.37	8.62	7.07	5.07	5.49
π	9.52	15.23	41.80**	28.67*	39.61**	80.07**
EXCH	17.33	28.68*	11.70	5.62	25.65*	14.73
CPBT	6.71	10.70	13.53	4.53	25.60*	13.98
TBILL	35.51**	20.58	24.80*	34.64**	11.72	12.03
DIV_{-1}	29.67**	49.37**	79.22**	76.68**	59.71**	58.44**

Note. Entries on first five rows are χ^2 with fourteen degrees of freedom; entries for DIV_{-1} are χ^2 with seven degrees of freedom. *denotes significance at the 5 percent level; **denotes significance at the 1 percent level.

in each group. These test statistics must be compared with the critical value of the χ^2 distribution with fourteen degrees of freedom. Although every variable (with the exception of real GNP) is significant for at least one group of companies, overall the statistics are low, showing that the explanatory power of these variables is limited.

After running these regressions, we test whether the resulting residual covariance matrix for each group of companies is diagonal. This is equivalent to applying the χ^2 test discussed above in the context of earnings to the correlation matrix of residuals. The χ^2 statistics for the test of a diagonal residual covariance matrix are equal to 146.38, 172.82, 185.11, 141.85, 209.88, and 247.62 for groups A, B, C, D, E, and F, respectively. These are all significant at the 1 percent level, so we easily reject the hypothesis that the $\epsilon_{i,t}$'s are uncorrelated across firms. In fact, these statistics are nearly the same as those for the raw returns. For groups B and C the χ^2 statistics for the OLS residuals are actually larger. Adding explanatory variables reduces the unexplained variance of returns for each company, but can lead to smaller reductions in covariances so that the correlations rise.

At this point, it is worth asking whether these correlations of returns residuals bear much relationship to the correlations of earnings. Ideally, one would want to know whether those companies whose returns are more correlated have expected present discounted values of earnings that are more correlated. Since we cannot measure expected present values, we check whether correlations of returns are related to the correlations of earnings changes displayed in Table II. To do this, we run regressions with the dependent variable,

$$c_{i,j} = \tan(\pi\rho_{ij}/2), \tag{6}$$

where ρ_{ij} is the residual correlation of returns for firms i and j. This transformation of the correlation has a range of minus to plus infinity. The same transformation applied to the earnings correlations yields $c^e_{i,j}$.

We use all of our two-firm correlations so that we have 126 observations (21 correlations for each group). The correlations within each group are not completely independent because the correlation matrix must be positive definite; we ignore this so our standard errors may be understated. Regressing $c_{i,j}$ on $c^e_{i,j}$ gives

$$\begin{aligned} c_{i,j} = \underset{(0.031)}{0.777} - \underset{(0.142)}{0.176}c^e_{i,j} \qquad R^2 = 0.012, \quad \text{NOB} = 126. \end{aligned} \tag{7}$$

Observe that there is essentially no connection between the two correlations. If anything, the point estimate suggests that firms whose earnings are more positively correlated have slightly more negatively correlated returns. Hence one must look elsewhere to explain the returns correlations.

We tested several variations on the regressions explaining returns with aggregate variables. First, we included the lagged dividend yield on the S&P 500 index as an explanatory variable. This variable has been shown by Fama and French [1988] and others to be a predictor of overall returns. Although it has no role in the theory, we add it because variables that predict expected returns are likely to predict comovements in returns. As Table III shows, this variable is a significant explanator of returns for all six groups of companies, so it might help explain the correlation of individual returns. Indeed, when it is included as a regressor, the χ^2 statistics for diagonal residual covariance matrices fall to 139.91, 158.42, 156.01, 119.46, 189.42, and 223.43. However, these are all still highly significant. As an alternative to the dividend yield, we added the lagged difference between the six-month commercial paper rate and the three-month Treasury bill rate, another predictor of stock returns. The resulting χ^2 statistics are 131.60, 163.77, 169.80, 126.14, 198.63, and 234.65. This variable reduces the χ^2's, but less so than the dividend yield.

One might argue that consumption is more likely than GNP to broadly affect expected returns. (We chose GNP because it is the underlying determinant of earnings.) Hence we also ran these regressions with the log change of real consumption of services and nondurables, current and lagged, as additional variables. The results are little changed; the χ^2 statistics are 141.57, 159.79, 196.24, 144.69, 190.90, and 250.17, respectively. Adding long-term interest rates and the log change of oil prices also makes little difference.

Next, we repeat these regressions including the market return. We do this for three reasons. First, as shown by Roll [1988] and Cutler, Poterba, and Summer [1989], macroeconomic variables explain very little of the movements in market indices, so regressions that exclude the market are likely to have correlated residuals. Second, the market as a whole may incorporate information about future macroeconomic conditions beyond what is present in the other macroeconomic variables.[5] Third, it is of interest to know

5. See Fama [1981] and Merton and Fischer [1984].

whether all the correlations in the returns are due to comovement with the market.

When we included the quarterly log change of the S&P 500 index, the test statistics dropped to 66.45, 86.59, 91.26, 91.74, 92.34, and 139.48. While the drop is substantial, these are still highly significant. Clearly, the correlation of stock returns for unrelated companies cannot be fully explained in terms of their correlation with broad indices. While the correlations are still significant, their average value is now close to zero. The reason is that any systematically positive correlation in returns will show up as a correlation with the mean of all returns, namely with the market return.

An obvious response to this finding is that the market captures some of the information people have about future macroeconomic conditions but not all the information. For instance, people might know something about future GNP and something about future inflation. Insofar as these two bits of information affect different firms differently, the correlations of returns are not completely summarized by the correlations of the returns with the market as a whole. Thus, additional information can explain comovement.[6] In the next section we use latent variables to test this possibility.

VI. Latent Variable Models

Investors have access to a wealth of information about future macroeconomic conditions, much of it qualitative in nature, and this should affect stock prices. Chen, Roll, and Ross [1986] anticipated as much when they included leads of macroeconomic variables as regressors. The problem with including leads is that investors cannot know the future with certainty, so that these leads are error-ridden measures of expectations. Unfortunately, the introduction of common explanatory variables subject to measurement error will itself lead to correlation among the residuals; the residual in every equation explaining returns is affected by the common measurement error.

Instead of using leads of macroeconomic variables, we use latent variables that represent their unobserved market expectations. The model has the following structure. Our regression equation is replaced by one with two types of explanatory variables:

$$Q_{i,t} = M'_t\gamma_{1,i} + \eta'_t\gamma_{2,i} + \epsilon_{i,t}, \tag{8}$$

6. This hypothesis is also considered by Shiller [1989]. Instead of using latent variables, he uses the difference between the stock price and the realized present discounted value of dividends to gauge the empirical significance of the hypothesis.

where M_t is a vector of observable macroeconomic variables, η_t is a vector of latent variables at t, and $\gamma_{1,i}$ and $\gamma_{2,i}$ are vectors of fixed coefficients.

For the time being, we let η_t include market expectations of certain macroeconomic variables at time $t + 1$, which we denote by Y_{t+1}. (The analysis would be unchanged if instead of representing expectations of $t + 1$ realizations, the latent variables were expectations of realizations further in the future.) Two additional equations are needed to ensure that the η's can be given this interpretation. The first is

$$Y_{t+1} = \eta_t + u_t, \tag{9}$$

which simply states that the future realizations of each macroeconomic variable equals the corresponding latent variable plus a forecast error.

The second equation is an econometric model for predicting future Y's:

$$Y_{t+1} = M'_t\alpha_0 + I'_t\alpha_1 + \epsilon_{Y,t+1}. \tag{10}$$

In equation (10) future values of the Y's are predicted using current values of a vector of macroeconomic variables M, as well as the values of a vector of additional variables (or instruments), I. The hypothesis of rational expectations and the requirement that agents observe the variables in I_t implies that the residuals u_t are uncorrelated with both M_t and I_t. The combination of (9) and (10) imposes economic structure on the latent variables.

Several comments about this latent variable procedure are in order. First, the instruments I_t include any observable variables useful for forecasting. It is thus natural to include broad market indices as instruments. The role of these indices in explaining individual returns then comes from their ability to predict macroeconomic variables.

Once a market index is included as an instrument, we have a model that cannot be rejected on the basis of the volatility of broad indices of stock prices alone. Thus, ours is not a test of excess volatility. This can be seen by considering a related model with only one equation such as (8), and the market return S_t is on the left-hand side:

$$S_t = M_t\gamma_{1,S} + \eta'_t\gamma_{2,S} + \epsilon_{S,t}. \tag{8'}$$

Equations like this, but without the latent variables η, were estimated by Roll [1988] and Cutler, Poterba, and Summers [1988], who obtained very low R^2's. Suppose that there is a single η

in (6′), a single Y in (9) and (10), and the only instrument is the market return S_t. If this model is estimated by maximum likelihood under the assumption that the residuals are normal, (6′) will fit perfectly, so there will be no excess comovement. The estimate of η_t will equal $M_t\alpha_0 + S_t\alpha_1$, where, in small samples, α_1 will differ from zero with probability one. Then, $\gamma_{2,S}$ will equal $1/\alpha_1$ while $\gamma_{1,S}$ will equal $-\alpha_0$ so that $\epsilon_{S,t}$ is zero.

Similarly, if returns move together only because of their correlation with a broad index, then versions of the model that have the index as the only instrument will have uncorrelated residuals. In this case, the model will easily explain the correlation of returns if many η's are included in (9) and if the individual returns are used as instruments.

We estimate equations (8), (9), and (10) for the seven returns in each group, using two latent variables: next quarter's expected real GNP, and next quarter's expected rate of growth of the GNP deflator. Expected inflation is the single determinant of the expected real interest rate on bonds once their nominal return is included as a regressor; this expected real interest rate is in turn related to discount rates and expected future economywide profits. We include in the vector M_t of observable variables current and lagged values of the growth rate of the GNP deflator (π), the growth rate of real GNP (Y), corporate profits before tax dividend by nominal GNP, (CPBT), the three-month Treasury bill rate (TBILL), and the exchange value of the dollar (EXCH). Finally, the vector I_t of instruments includes the current and lagged values of the S&P 500 index normalized by nominal GNP (S), the log first difference of the monetary base (BASE), and the growth rate of the real price of crude oil (CRUDE). Since we have already seen that these variables are not significantly related to earnings once one controls for GNP, it is plausible to assume that they affect stock returns only through their ability to predict aggregate economic activity.

We estimate this system via maximum likelihood, assuming that all of the residuals are normal, with and without the constraint of a diagonal variance-covariance matrix for the $\epsilon_{i,t}$'s. We then perform a likelihood ratio test of the diagonality restrictions. Note that even the model that does not impose a diagonal covariance matrix imposes some constraints; we use more instruments than latent variables so that the system is overidentified. Estimation is done using LISREL.[7] Besides yielding parameter

7. See Joreskog and Sorbom [1986] for an introduction to LISREL.

estimates, LISREL computes the value of the likelihood function, making likelihood ratio tests straightforward.

Parameter estimates for the constrained system are shown in Table IV for group A. (Parameter estimates for the other five groups are similar and are omitted to save space.) Consistent with the results in the previous section, the observable macroeconomic variables other than the Treasury bill rate are not significant as explanators of returns. On the other hand, the latent variable for expected real GNP growth is almost uniformly significant for all companies in all six groups. Also, the two latent variables together account for much of the correlation of returns that was left unexplained by the simple regression models. Both latent variables have estimated coefficients that are positive, as expected; higher future GNP raises earnings, and higher future inflation, for a given nominal interest rate, lowers discount rates. It may seem peculiar that expected future output depends positively on current crude oil prices and negatively on the monetary base. Indeed, in our sample the raw correlation between future output growth and current oil price changes is negative, while that with the base is positive. But, the conditional correlation, once one controls for the market return and the other included variables, can easily have the opposite sign.

The χ^2 statistics for a test of a diagonal covariance matrix for the $\epsilon_{i,t}$'s are 54.97, 48.76, 33.78, 37.43, 42.12, and 45.41, respectively, for the six groups. These statistics are much lower than those calculated for the residuals of the regression models. However, they are significant at the 1 percent level for groups A, B, D, E, and F, and at the 5 percent level for group C. Thus, there is still excess comovement in returns.

We also tested the overidentifying restrictions imposed by the models whose diagonal matrix is unconstrained. These restrictions are due to our use of a large number of instruments for the latent variables. There are 39 of these restrictions. (The unconstrained model has 42 correlations between the 6 lagged and unlagged instruments and the 7 returns, 12 correlations between the instruments and the two macroeconomic variables, and 14 correlations between the macroeconomic variables and the returns, for a total of 68 parameters. The constrained model has 14 correlations between the latent variables and the returns, 12 correlations between the instruments and the latent variables, and 3 parameters of the variance covariance matrix of the latent variables for a total of 29 parameters.) The test statistics, distributed as χ^2 (39), are 46.05, 61.07, 60.48, 63.88, 59.43, and 61.54 for groups A through F. The statistic for group A is below the 5 percent critical value, but the others are not, suggesting that these models are somewhat misspecified.

We also estimated alternative versions of these latent variable models to check the robustness of our results. First, one might argue that expectations one quarter ahead is too short a horizon. To test this, we let the latent variables represent expected real GNP growth and inflation two quarters ahead. The results are similar; the likelihood ratio statistics for the test of a diagonal covariance matrix are 44.56, 48.77, 31.05, and 33.58, respectively, for groups A, B, C, and D, which are all significant at the 5 percent level.[8]

Second, we checked whether our results are sensitive to the Crash of 1987, which one might view as an outlier. To do this, we reestimated the model using data for the third quarter of 1969 to the third quarter of 1987. The resulting χ^2 statistics for groups A, B, C, D, and E are 53.83, 47.12, 34.12, 34.87, and 41.8, respectively, which are all significant at the 1 percent level. (Group F did not converge.) These numbers are close to those for the longer time horizon, presumably because the market return is included as an instrument.[9]

Finally, we formed an additional group of 7 companies (Group G) from our original 42 companies. This group was chosen so that the correlations of normalized earnings changes are as low as possible, and none are statistically significant at the 5 percent level. (Groups A to F each have two or three significant correlations, although the correlations are groupwise insignificant.) The companies in this new group are Delta Airlines (DAL), Giant Group Ltd. (GG), International Paper (IP), Chevron (CHV), Newmont Mining (NEM), Stone and Webster (SW), and Melville Corp. (MEL). Part A of Table V shows the earnings correlations; note that the χ^2 statistic for groupwise significance is only 10.23. Nonetheless, as Table V shows, the residual correlations from the OLS regressions of returns and from the latent variable model remain highly significant. The likelihood ratio statistic for a diagonal covariance matrix in the latent variable model is 59.36, which is significant at 1 percent. Thus, even when companies are grouped to minimize the

8. We were unable to obtain convergence when estimating the model for groups E and F. We also tried to estimate a model in which the latent variables represent expectations four quarters ahead. This converged only for group B. The test statistic in that case was 48.58, which is significant at the 1 percent level.

9. We also tested the explanatory power of the lagged dividend yield by including it as an additional explanatory variable. The χ^2 statistic for group B is 55.57, which is higher than when this variable is excluded. (We could not obtain convergence for the other groups.)

TABLE IV
LATENT VARIABLE MODEL: GROUP A

	IP	CHV	NEM	SW	CQ	KR	CHL	η_π	η_y
η_π	1.245	1.221	1.405	1.297	0.824	1.027	1.443		
	(1.6)	(1.5)	(1.7)	(1.8)	(1.1)	(1.7)	(1.8)		
η_y	2.336	2.323	2.386	1.986	2.108	1.700	2.303		
	(2.3)	(2.2)	(2.2)	(2.2)	(2.2)	(2.1)	(2.2)		
π	−0.533	−0.212	−0.384	−0.479	−0.341	−0.269	−0.612	0.523	−0.075
	(−1.0)	(−0.4)	(−0.7)	(−1.0)	(−0.7)	(−0.7)	(−1.2)	(5.2)	(−0.5)
$\pi(-1)$	0.224	−0.001	0.011	0.104	0.303	0.026	0.195	0.322	−0.211
	(0.5)	(−0.0)	(0.0)	(0.3)	(0.8)	(0.1)	(0.5)	(3.2)	(−1.4)
Y	0.461	0.460	0.541	0.392	0.176	0.003	0.195	−0.132	−0.048
	(1.1)	(1.1)	(1.2)	(1.0)	(0.5)	(0.0)	(0.5)	(−1.3)	(−0.3)
$Y(-1)$	0.071	0.032	0.080	−0.144	−0.063	−0.007	0.137	−0.243	0.138
	(0.2)	(0.1)	(0.2)	(−0.5)	(−0.2)	(−0.0)	(0.4)	(−2.8)	(1.1)
CPBT	−0.756	−0.645	−1.104	−0.747	−0.531	−0.298	−0.603	−0.031	0.251
	(−0.9)	(−0.7)	(−1.2)	(−1.0)	(−0.7)	(−0.4)	(−0.7)	(−0.1)	(0.8)
CPBT(−1)	0.351	0.380	0.811	0.469	0.388	0.339	0.460	0.240	−0.294
	(0.5)	(0.5)	(1.1)	(0.8)	(0.6)	(0.6)	(0.6)	(1.4)	(−1.1)
TBILL	−1.497	−1.195	−1.709	−1.499	−1.171	−1.634	−1.685	0.336	0.356
	(−1.6)	(−1.3)	(−1.8)	(−1.8)	(−1.4)	(−2.3)	(−1.8)	(1.7)	(1.2)
TBILL(−1)	2.483	2.126	2.728	2.398	2.071	2.504	2.776	−0.324	−0.800
	(2.0)	(1.7)	(2.2)	(2.2)	(1.8)	(2.6)	(2.3)	(−1.5)	(−2.4)
EXCH	−0.329	−0.410	−0.198	−0.122	−0.338	−0.166	−0.114	0.022	0.063
	(−1.0)	(−1.3)	(−0.6)	(−0.4)	(−1.1)	(−0.7)	(−0.4)	(0.3)	(0.5)
EXCH(−1)	−0.223	−0.103	−0.318	−0.504	−0.062	−0.218	−0.251	0.192	−0.039
	(−0.7)	(−0.3)	(−1.0)	(−1.8)	(−0.2)	(−0.9)	(−0.8)	(2.5)	(−0.3)
CRUDE								−0.248	0.200
								(−3.7)	(2.6)

TABLE IV
(CONTINUED)

	IP	CHV	NEM	SW	CQ	KR	CHL	η_π	η_y
CRUDE(−1)								0.248	−0.169
								(4.0)	(−2.4)
BASE								0.179	−0.143
								(2.9)	(−2.3)
BASE(−1)								0.245	−0.169
								(3.5)	(−2.3)
S								0.561	0.692
								(1.9)	(1.6)
$S(-1)$								−0.305	−0.920
								(−1.0)	(−2.1)
R^2	0.49	0.42	0.43	0.48	0.36	0.42	0.49		

$\chi^2(21) = 54.97$

correlation of their earnings, we still cannot account for the comovement of their returns.

VII. Explaining Comovement with Firm and Investor Characteristics

If it is accepted that expectations of future macroeconomic conditions do not account for comovement of returns, then what does? One possibility is that sets of securities are predominantly held by certain groups of investors and that these groups are affected by variables which have little to do with the aggregate economy. These variables can be related to the future income of the group (and thus to the discount rates that the group applies to future earnings), or they can be variables that change the group's perceptions of future earnings. Changes in perceptions may be rational, or might represent irrational herd behavior. Lee, Shleifer, and Thaler [1991] also test whether the prices of securities held by certain investors move together. They show that prices of stocks with low institutional holdings, and stocks of small companies, move together with the discounts on closed-end funds (which are also predominantly held by individuals).

We first examine whether stocks in a particular size class tend to move together. This was considered by Huberman, Kandel, and Karolyi [1988], who found that stock returns for firms in different size groups move together, even after controlling for industry-based comovements. We show that size explains comovement in our sample as well, so our results are consistent with theirs. This finding does not demonstrate that market segmentation is causing comovement. Indeed, they suggest that small stocks may move together because they are more affected by risk premiums, which in our setting would mean that a missing macroeconomic factor explains stock returns of small companies. On the other hand, comovement among stocks of equal size may indicate market segmentation because certain investors, namely institutions, hold disproportionate shares of their portfolios in the stocks of large companies [Arbel, 1985].

We try to explain excess (i.e., residual) correlations among returns by variables accounting for similarity in size and similarity of institutional ownership. Our limited sample size precludes a nonparametric analysis of the relationship between our correlations and size or institutional ownership. We thus proceed para-

TABLE V

Part A

Group G: Correlations of Normalized Earnings Changes

	DAL	GG	IP	CHV	NEM	SW	MES
DAL	1.000						
GG	0.059	1.000					
IP	−0.137	−0.098	1.000				
CHV	0.042	−0.141	0.007	1.000			
NEM	−0.048	−0.088	0.044	−0.105	1.000		
SW	0.102	0.080	−0.141	−0.125	0.028	1.000	
MES	−0.022	−0.006	−0.076	−0.021	−0.060	0.054	1.000

$\chi^2(21) = 10.23$

Part B

Group G: Residual correlations from OLS regressions of returns

	DAL	GG	IP	CHV	NEM	SW	MES
DAL	1.000						
GG	0.290	1.000					
IP	0.350	0.506	1.000				
CHV	0.355	0.588	0.443	1.000			
NEM	0.197	0.510	0.344	0.389	1.000		
SW	0.253	0.432	0.236	0.610	0.332	1.000	
MES	0.516	0.238	0.269	0.392	0.083	0.344	1.000

$\chi^2(21) = 165.47$

Part C

Group G: Residual correlations from latent variable model of returns

	DAL	GG	IP	CHV	NEM	SW	MES
DAL	1.000						
GG	−0.160	1.000					
IP	0.056	0.147	1.000				
CHV	−0.069	0.059	−0.001	1.000			
NEM	−0.135	−0.195	−0.023	−0.061	1.000		
SW	−0.098	−0.062	−0.227	0.181	−0.035	1.000	
MES	0.380	−0.158	0.004	0.084	−0.250	−0.082	1.000

$\chi^2(21) = 59.36$

metrically by analyzing the relationship between residual correlations in returns and several size and ownership variables.

We compute three size variables based on the market value of the firm's equity. We constructed analogous measures using value added, but these did not explain the correlation of returns. First, we compute for each company i the equity capitalization at the end of the year as defined by the product of the share price and the number of shares outstanding. We then compute the average equity capitalization between 1971 and 1989 for each firm, which we denote by m_i. For each pair of firms $\{i,j\}$ in our sample, our three measures of size are

$$S^s_{i,j} = \frac{\min(m_i,m_j)}{\max(m_i,m_j)}$$
$$S^p_{i,j} = m_i m_j \tag{11}$$
$$S^m_{i,j} = \min(m_i,m_j).$$

The first is a measure of similarity; it is small when the firms have very different equity capitalizations, and reaches a maximum when their capitalizations are equal. The other two measures of size simply increase when the sizes of the component firms increase.

We obtained measures of institutional ownership from the Standard and Poor's Stock Guides. Because the definitions change in 1981, we only use data from 1981 to 1989. For each company in our sample we divided the number of shares held by institutions by total shares outstanding. We obtained these ratios for July of each year and then computed a 1981–1989 average for each firm, which we denote by n_i. We then constructed three measures of ownership, H^s, H^p, and H^m by applying (11) to these n_i's.

Table VI reports the results of regressions of the residual correlations of returns as transformed in (6) and the S and H variables. Part A is based on the residual correlations from the OLS regressions that exclude the market return. Part B is based on residual correlations from OLS regressions that control for the market return, and Part C is based on the residual correlations from our latent variable model.

Observe from the first three rows of each of part of Table VI that the S^s size variable is positively correlated with the residual correlations (significantly so in the case of the first OLS model). Hence our procedure reproduces the results of Huberman, Kandel, and Karloyi [1988], since stocks of similar market capitalization are more positively correlated than stocks of different market capitalization. The variables based on absolute size do not explain the residual correlations of returns; it is size similarity, not size per se that matters. The standard deviation of S^s is 0.28. The coefficient from Part A of Table VI then implies that a one-standard-deviation increase in S^s is associated with an increase in the correlation of returns of 0.06. To put this in perspective, the standard deviation of the correlation of returns is about 0.1. Thus, a one-standard-deviation change in S^s is associated with a change in the correlation equal to 60 percent of the standard deviation of the correlation.

Rows (4) through (6) show the individual effects of our institutional holdings variables. In the OLS case, H^p and H^m are strongly correlated with the residual correlations; pairs of stocks that are held disproportionately by institutions covary more. The correlation of the H^s variable with the latent variable residuals is weaker. Indeed, rows seven and eight show that, once one controls for either H^m or H^p, the H^s variable becomes insignificant, even for the OLS regressions that do not include the market. Thus, it is the size of institutional holdings itself rather than the similarity of institutional holdings that leads to comovement.

The standard deviation of H^p is 0.10. Using the coefficient in row (5) from Part A, a one-standard-deviation increase in H^p is thus associated with a 0.1 increase in the correlation of returns, i.e., a one-standard-deviation increase in the correlation. The standard deviation of H^m is 0.14, so, with a coefficient of 0.78, a one-standard-deviation increase in H^m is also associated with a one-standard-deviation increase in the correlation. Both of these are economically significant effects.

These results are similar to those of Lee, Shleifer, and Thaler [1991]. They find that stocks held by individuals move with the discounts on closed-end funds, and we find comovement among stocks that are held disproportionately by institutional investors. In both cases, common ownership is associated with comovement.

Perhaps the most interesting results are in the last two rows of Parts A and B, which show the effect of including both the statistically significant size variable S^s and the two more significant institutional holdings variables. The inclusion of our institutional holdings variable eliminates the explanatory power of the market capitalization size variable. Hence it appears that a good explanation for the size effect obtained by Huberman, Kandel, and Karolyi [1988] is the tendency of institutions to hold companies with large capitalizations.

The results based on the residuals of the latent variable model are qualitatively similar though the coefficients are less statistically significant. One reason is that the introduction of latent variables automatically reduces the cross correlations of residuals. (As we saw, this is not the case when variables are added to the OLS

TABLE VI
EXPLAINING THE RESIDUAL CORRELATIONS

Eq.	Cons.	H^s	H^p	H^m	S^s	S^p	S^m	R^2
				A. OLS residuals (without market)				
(1)	0.685				0.231			0.035
	(0.054)				(0.108)			
(2)	0.799					−7E-9		0.008
	(0.036)					(7E-9)		
(3)	0.772						1E-5	0.0004
	(0.047)						(5E-5)	
(4)	0.520	0.364						0.045
	(0.112)	(0.151)						
(5)	0.530		1.039					0.095
	(0.076)		(0.289)					
(6)	0.461			0.782				0.094
	(0.094)			(0.219)				
(7)	0.507	0.056	0.971					0.095
	(0.110)	(0.189)	(0.370)					
(8)	0.492		0.582	0.364				0.096
	(0.106)		(0.929)	(0.703)				
(9)	0.509	−0.218		1.049				0.099
	(0.110)	(0.260)		(0.386)				
(10)	0.521		0.941		0.081			0.098
	(0.077)		(0.322)		(0.117)			
(11)	0.460			0.711	0.076			0.097
	(0.094)			(0.246)	(0.118)			
				B. OLS residuals (with market)				
(1)	0.399				0.165			0.019
	(0.052)				(0.105)			
(2)	0.491					−8E-9		0.015
	(0.035)					(6E-9)		
(3)	0.487						−2E-5	0.003
	(0.046)						(4E-5)	
(4)	0.469	−0.003						4E-6
	(0.112)	(0.150)						
(5)	0.263		0.849					0.067
	(0.074)		(0.285)					
(6)	0.286			0.443				0.032
	(0.094)			(0.219)				

TABLE VI
(CONTINUED)

Eq.	Cons.	H^s	H^p	H^m	S^s	S^p	S^m	R^2
			B. OLS residuals (with market)					
(7)	0.450	−0.445	1.392					0.110
	(0.106)	(0.182)	(0.356)					
(8)	0.445		3.018	−1.728				0.114
	(0.102)		(0.893)	(0.676)				
(9)	0.454	−0.773		1.386				0.100
	(0.106)	(0.252)		(0.374)				
(10)	0.258		0.805		0.036			0.068
	(0.076)		(0.318)		(0.116)			
(11)	0.284			0.361	0.086			0.036
	(0.094)			(0.247)	(0.119)			
			C. Residuals of latent variable model					
(1)	−0.041				0.101			0.008
	(0.050)				(0.100)			
(2)	−0.009					3E-9		0.002
	(0.033)					(6E-9)		
(3)	−0.051						7E-5	0.020
	(0.043)						(4E-5)	
(4)	0.019	−0.026						0.0003
	(0.105)	(0.142)						
(5)	−0.061		0.255					0.007
	(0.072)		(0.277)					
(6)	−0.047			0.117				0.002
	(0.090)			(0.210)				
(7)	0.013	−0.175	0.468					0.014
	(0.105)	(0.180)	(0.353)					
(8)	0.012		1.117	−0.686				0.015
	(0.101)		(0.887)	(0.671)				
(9)	0.014	−0.283		0.462				0.013
	(0.105)	(0.249)		(0.369)				
(10)	−0.070		0.165		0.074			0.010
	(0.074)		(0.309)		(0.112)			
(11)	−0.049			0.027	0.095			0.008
	(0.090)			(0.236)	(0.113)			

regression.) Hence the remaining cross correlations are probably downward biased estimates of the true correlations.

VIII. CONCLUSIONS

We have argued that correlations among stock returns for companies in unrelated lines of business should be due to changes in current or expected future values of macroeconomic variables. Observable variables other than a broad market index explain little of the movements in individual returns, and even accounting for a market index leaves returns highly correlated. Adding latent variables that represent unobservable forecasts of economic conditions accounts for some of this correlation, but not all. For every variation of our model that we have tried, and for every group of companies that we have examined, we find excess correlation of returns. If investors' utilities are a function of a single consumption index (as most models of asset pricing posit), then our results conflict with the present value model of security valuation.

One might argue that our tests are incomplete, in that we may have excluded some important macroeconomic variables from our specifications. This possibility cannot be ruled out. This problem is analogous to that which arises in empirical tests of the APT; if a sufficiently large number of factors is included, the model cannot be rejected. The voluminous literature testing the APT thus restricts itself to using only a small number of factors and investigating whether these suffice to explain differences in ex ante returns across stocks. In the same spirit we have studied whether a reasonable number of macroeconomic variables and latent variables can explain the correlation of returns.

It is important to emphasize, however, that simply adding an additional latent variable need not reduce the χ^2 statistics for our tests of a diagonal covariance matrix. In fact, our work on commodity prices shows that adding statistically insignificant latent variables can increase these statistics.[10] In any case, more extensive experiments are needed to see whether other variables can better explain returns and their joint movements.[11]

10. In Pindyck and Rotemberg [1990] we tested for excess comovement of commodity prices by first estimating a model with one latent variable, and then with two. The corresponding χ^2 statistics generally increased.

11. The excluded variables, if there are any, represent either changes in discount rates or changes in expected future earnings. It might be possible to test the hypothesis that variables that explain discount rates have not been excluded by adapting the technique of Hansen and Jaganathan [1988]. Their method uses data on returns to measure the extent to which marginal rates of substitution (or discount rates) vary. It might thus prove useful in checking whether the correlation of returns is due to variations in discount rates.

We believe, however, that is it more promising to pursue the hypothesis that stock comovement is due to market segmentation. This hypothesis is consistent with our results as well as those of Lee, Shleifer, and Thaler [1991], who also found that the prices of stocks held in common move together. An unresolved question is what mechanism leads to covariation in the prices of the securities held by particular groups. One possibility is that different groups have different sentiments about the future payoffs of different securities. This would explain why different groups hold different portfolios in the first place. Thus, the existence of groupwide sentiments can explain both comovement and the market segmentation with which it associated.

Another possibility is that segmented ownership is the result of differences in transactions costs. Then, changes in the expected future income of a group would affect that group's discount rate and thus the prices of the securities that it holds. The problem with this explanation is that transactions costs do not, in fact, appear to vary across groups in the required way.

SLOAN SCHOOL, MASSACHUSETTS INSTITUTE OF TECHNOLOGY

REFERENCES

Arbel, A., "Generic Stocks: An Old Product in a New Package," *Journal of Portfolio Management,* XI (1985), 4–13.
Burmeister, Edwin, and Marjorie B. McElroy, "Joint Estimation of Factor Sensitivities and Risk Premia for the Arbitrage Pricing Theory," *Journal of Finance,* XLIII (1988), 721–33.
Chen, Nai-Fu, Richard Roll, and Stephen A. Ross, "Economic Forces and the Stock Market," *Journal of Business,* LIX (1986), 383–403.
Cutler, David M., James M. Poterba, and Lawrence H. Summers, "What Moves Stock Prices?" *Journal of Portfolio Management,* XV (1989), 4–12.
Fama, Eugene F., "Stock Returns, Real Activity, Inflation, and Money," *American Economic Review,* LXXI (1981), 545–65.
Fama, Eugene F., and Kenneth R. French, "Dividend Yields and Expected Stock Returns," *Journal of Financial Economics,* XXII (1988), 3–25.
Fischer, Stanley, and Robert C. Merton, "Macroeconomics and Finance: The Role of the Stock Market," in *Essays on Macroeconomic Implications of Financial and Labor Markets and Political Processes,* K. Brunner and A. Meltzer, eds., Carnegie-Rochester Conference Series on Public Policy, XXI (1984), 57–108.
Hansen, Lars Peter, and Ravi Jaganathan, "Restrictions on Intertemporal Marginal Rates of Substitution Implied by Asset Prices," mimeo, November 1988.
Huberman, Gur, Shmuel Kandel, and G. Andrew Karolyi, "Size and Industry Related Covariations of Stock Returns," University of Alberta, Working Paper 991/88, November 1988.
Joreskog, Karl G., and Dag Sorbom, "LISREL User's Guide," 1986.
King, Benjamin F., "Market and Industry Factors in Stock Price Behavior," *Journal of Business,* XXXIX (1966), 139–90.
King, Mervyn, and Sushil Wadhwani, "Transmission of Volatility Between Stock Markets," NBER Working Paper No. 2910, March 1989.
Lee, Charles M. C., Andrei Shleifer, and Richard H. Thaler, "Investor Sentiment and the Closed-End Fund Puzzle," *Journal of Finance,* XLVI (1991), 75–109.
Lee, Cheng F., and Joseph D. Vinso, "Single vs. Simultaneous Equation Models in Capital Asset Pricing: The Role of Firm-Related Variables," *Journal of Business Research,* VIII (1980), 65–80.

LeRoy, Stephen F., and Richard D. Porter, "The Present Value Relation: Tests Based on Implied Variance Bounds," *Econometrica,* XLIX (1981), 555–74.
Mankiw, N. G., "The Equity Premium and the Concentration of Aggregate Shocks," *Journal of Financial Economics,* XVII (1986), 211–19.
McElroy, Marjorie B., and Edwin Burmeister, "Arbitrage Pricing Theory as a Restricted Nonlinear Multivariate Regression Model," *Journal of Business & Economic Statistics,* VI (1988), 29–42.
Meyers, Stephen L., "A Re-Examination of Market and Industry Factors in Stock Price Behavior," *Journal of Finance,* XXVIII (1973), 695–706.
Morrison, Donald, *Multivariate Statistics Methods* (New York: McGraw-Hill, 1967).
Pindyck, Robert S., and Julio J. Rotemberg, "The Excess Co-Movement of Commodity Prices," *Economic Journal,* C (1990), 1173–89.
Roll, Richard, "R^2," *Journal of Finance,* XLIII (1988), 541–66.
Shiller, Robert J., "Do Stock Prices Move Too Much to Be Justified by Subsequent Changes in Dividends?" *American Economic Review,* LXXI (1981), 421–36.
——, "Comovements in Stock Prices and Comovements in Dividends," *Journal of Finance,* XLIV (1989), 719–29.