Hui Guo
Federal Reserve Bank of St. Louis

On the Out-of-Sample Predictability of Stock Market Returns*

In this paper, I provide new evidence of the out-of-sample predictability of stock returns. In particular, I find that the consumption-wealth ratio in conjunction with a measure of aggregate stock market volatility exhibits substantial out-of-sample forecasting power for excess stock market returns. Also, simple trading strategies based on the documented predictability generate returns of higher mean and lower volatility than the buy-and-hold strategy does, and this difference is economically important.

There is an ongoing debate about stock return predictability in time-series data. Campbell (1987) and Fama and French (1989), among many others, find that macro variables such as the dividend yield, the default premium, the term premium, and the short-term interest rate forecast excess stock market returns. However, Bossaerts and Hillion (1999), Ang and Bekaert (2001), and Goyal and Welch (2003) cast doubt on the in-sample evidence documented by the early authors by showing that these variables have negligible out-of-sample predictive power.

In this paper, I provide new evidence of the out-of-sample predictability of stock returns. In particular, I find that the consumption-wealth ratio (*cay*) by Lettau and Ludvigson (2001)—the error term from the cointegration relation among consumption, wealth, and labor income—exhibits substantial out-of-sample forecasting abilities for stock returns if augmented by a measure of aggregate stock market volatility (σ_m^2). More important, the improvement of the forecast model of *cay* augmented by σ_m^2 over the model of *cay* by itself is statistically significant. My results reflect a classic omitted-variable problem: While *cay* and

* I want to thank Martin Lettau, Sydney Ludvigson, Mike Pakko, Albert Madansky (the editor), and an anonymous referee for very helpful suggestions. George Fortier provided excellent editorial support. The views expressed in this paper are those of the author and do not necessarily reflect the official positions of the Federal Reserve Bank of St. Louis or the Federal Reserve System. Contact the author at hui.guo@stls.frb.org.

[*Journal of Business*, 2006, vol. 79, no. 2]

0021-9398/2006/7902-0007$10.00

σ_m^2 are negatively related to one another, they are both positively correlated with future stock returns.[1]

For robustness, I also investigate whether one can use simple trading strategies to exploit the predictability documented in this paper. As suggested by Leitch and Tanner (1991), this evaluation criterion is potentially more sensible than the statistical counterpart. I consider two widely used and relatively naive portfolio strategies. First, following Breen, Glosten, and Jagannathan (1989), among others, one holds stocks if the predicted excess return is positive and hold bonds otherwise. In the second strategy, which has been used by Johannes, Polson, and Stroud (2002), among others, I allocate wealth between stocks and bonds according to the formula of the static capital asset pricing model (CAPM). I find that the managed portfolio generates higher mean returns with lower volatility than the market portfolio, and this difference is economically important. For example, the certainty equivalence calculation suggests that an investor would agree to pay annual fees of 2%–3% to hold the managed portfolio rather than the market portfolio over the period 1968: Q2–2002:Q4. Also, neither the CAPM nor the Fama and French (1993) three-factor model can explain returns on the managed portfolio, and I reject the null hypothesis of no market timing ability using Cumby and Modest's (1987) test. Moreover, my trading strategies require relatively infrequent rebalancing of portfolios, and therefore, these results are robust to the adjustment of reasonable transaction costs. Interestingly, consistent with Pesaran and Timmermann (1995), I find substantial variations in the profitability of trading strategies through time.

My results are in sharp contrast with those of Bossaerts and Hillion (1999), Ang and Bekaert (2001), and Goyal and Welch (2003), as mentioned above. This difference is explained by the fact that my forecasting variables drive out most variables used by the early authors, including the dividend yield, the default premium, and the term premium. There is one exception. The stochastically detrended risk-free rate (*rrel*) used by Campbell, Lo, and MacKinlay (1997), among others, provides information beyond *cay* and σ_m^2 about future stock returns in the in-sample regression over the post–World War II period, although it becomes insignificant after 1980.[2] I also find mixed evidence of its out-of-sample forecast performance.

My forecasting variables are motivated by those in the paper by Guo (2004),

1. Brennan and Xia (2002) argue that the forecasting power of *cay* is spurious because if calendar time is used in place of consumption, the resulting cointegration error, *tay*, performs as well as or better than *cay* in predicting stock returns. In the Appendix, I show that *cay* always drives out *tay* if one adds past stock market variance and the stochastically detrended risk-free rate to the forecasting equation. Therefore, although the results by Brennan and Xia are interesting because they reflect an unstable relation between *cay* and excess stock market returns due to the omitted-variable problem documented in this paper, they do not pose a challenge to the forecasting power of *cay*.

2. The short-term interest rate and stock prices fell dramatically in 2001–2. This episode has a large impact on the forecasting power of *rrel*: It is significant if these two years are excluded from the post-1980 sample.

who shows that, in addition to the risk premium as stressed by standard models, investors also require a liquidity premium on stocks because of limited stock market participation. Therefore, σ_m^2 and *cay* forecast stock returns because they proxy for the risk and liquidity premiums, respectively.[3] Moreover, Guo shows that, although the two variables are both positively related to future stock returns, they could be negatively correlated with one another, as observed in the data.

The paper is organized as follows. I discuss the data in Section I and report the out-of-sample forecasting exercises in Section II. Some simple trading strategies are analyzed in Section III, and Section IV offers some concluding remarks.

I. Data

The consumption, net worth, labor income data, and the generated variable *cay* over the period 1952:Q2–2002:Q3 are obtained from Martin Lettau at New York University. I use the value-weighted stock market return obtained from the Center for Research in Security Prices (CRSP) as a measure of market returns. The risk-free rate obtained from CRSP is used to construct excess stock returns. As in Merton (1980) and many others, I construct the realized stock market variance, σ_m^2, using the daily stock market return data, which are obtained from Schwert (1990) before July 1962 and from CRSP thereafter. Following Campbell et al. (2001), I adjust downward the realized stock market variance for 1987:Q4, on which the 1987 stock market crash has confounding effects. The stochastically detrended risk-free rate, *rrel*, is the difference between the nominal risk-free rate and its last four-quarter average.

Table 1, which includes the full sample and two subsample periods, presents summary statistics of excess stock market return, $r_m - r_f$, and its forecasting variables used in this paper. It should be noted that the autocorrelation coefficients of the forecasting variables are less than 0.90 in both the full sample and the subsamples. There are some differences between the two subsamples. First, $\widehat{cay}$ is more negatively related with σ_m^2 in the second half (panel C) than the first half (panel B) of the sample. Second, while $\widehat{cay}$ is negatively related to *rrel* in panel B, the two are slightly positively related in panel C. Third, excess stock market return, $r_m - r_f$, is more negatively related with *rrel* in panel B than in panel C.

Figures 1–3 plot the forecasting variables through time. While $\widehat{cay}$ (fig. 1) fell sharply, σ_m^2 (fig. 2) rose dramatically during the second half of the 1990s. This pattern explains the strong negative relation between the two variables as reported in table 1. Also, *rrel* (fig. 3) fell steeply during the stock market

3. Patelis (1997) suggests that variables such as *rrel* reflect the stance of monetary policies, which have state-dependent effects on real economic activities through a credit channel (e.g., Bernanke and Gertler 1989).

TABLE 1 Summary Statistics

	$r_m - r_f$ (1)	$\widehat{cay}$ (2)	σ_m^2 (3)	rel (4)
	A. 1952:Q2–2002:Q4 Correlation Matrix			
$r_m - r_f$	1.000			
$\widehat{cay}$	.334	1.000		
σ_m^2	−.415	−.392	1.000	
rrel	−.260	−.138	−.087	1.000
	Univariate Statistics			
Mean	.017	.000	.004	.000
Standard deviation	.084	.013	.004	.003
Autocorrelation	.070	.831	.488	.720
	B. 1952:Q2–1977:Q4 Correlation Matrix			
$r_m - r_f$	1.000			
$\widehat{cay}$	.403	1.000		
σ_m^2	−.466	−.104	1.000	
rrel	−.397	−.507	.075	1.000
	Univariate Statistics			
Mean	.016	.000	.003	.000
Standard deviation	.083	.010	.003	.002
Autocorrelation	.154	.779	.391	.743
	C. 1978:Q1–2002:Q4 Correlation Matrix			
$r_m - r_f$	1.000			
$\widehat{cay}$	.293	1.000		
σ_m^2	−.424	−.633	1.000	
rrel	−.185	.059	−.127	1.000
	Univariate Statistics			
Mean	.017	.000	.005	.000
Standard deviation	.086	.015	.005	.003
Autocorrelation	−.012	.854	.481	.709

NOTE.—$r_m - r_f$ is the excess stock market return. The consumption-wealth ratio, *cay*, is the error term from the cointegration relation among consumption, labor income, and net worth. Realized stock market variance, σ_m^2, is constructed using daily data as in Merton (1980). The stochastically detrended risk-free rate, *rrel*, is the difference between a nominal risk-free rate and its last four-quarter average.

"bubble" burst in 2001–2. As I show below, this episode weakens the forecasting ability of σ_m^2 and *rrel* for stock market returns. However, the stock market correction in 2001–2 reinforces the forecasting ability of $\widehat{cay}$, which has been below its historical average since 1997. Nevertheless, my main results are not sensitive to whether I include these two years in the sample.

I first discuss the in-sample regression results. As argued by Inoue and Kilian (2002), while out-of-sample tests are not necessarily more reliable than in-sample tests, in-sample tests are more powerful than out-of-sample tests, even asymptotically. Table 2 presents the ordinary least squares estimation results, with heteroskedasticity- and autocorrelation-corrected *t*-statistics in

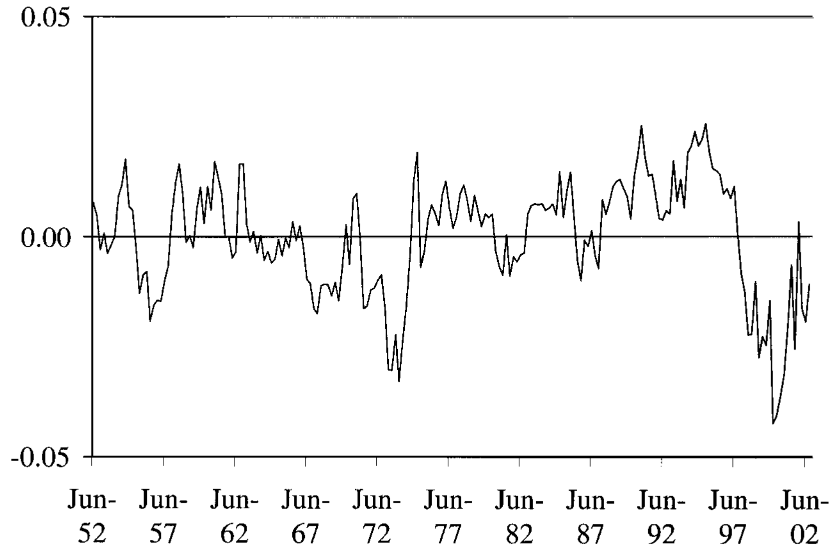

FIG. 1.—Consumption-wealth ratio

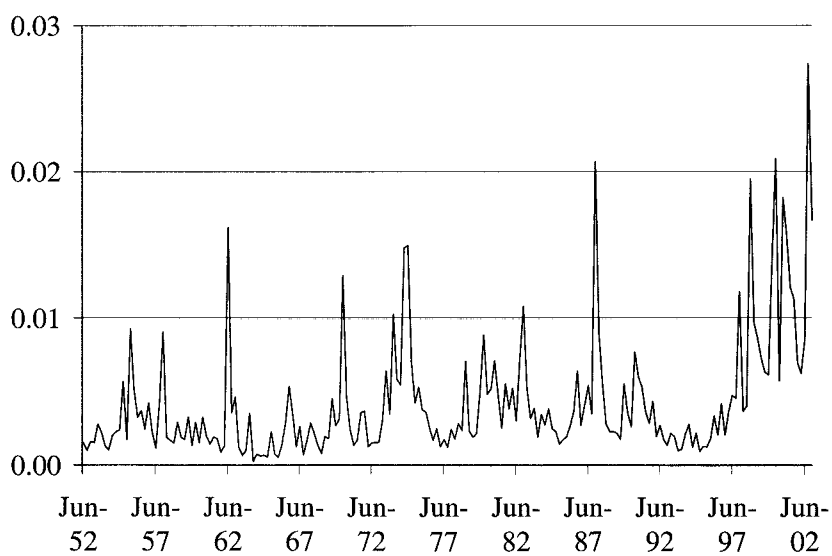

FIG. 2.—Realized stock market variance

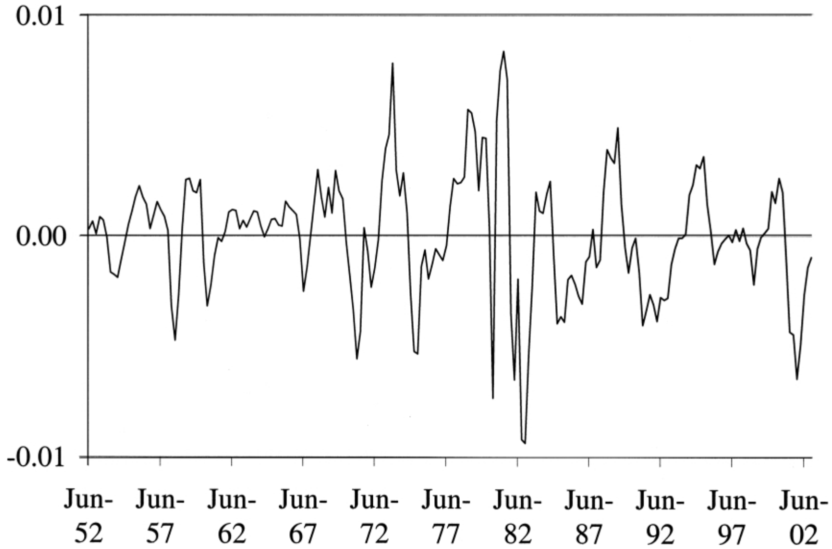

FIG. 3.—Stochastically detrended risk-free rate

parentheses. It should be noted that I construct $\widehat{cay}$ using the full sample, even in the subsample analysis.

Panel A is the full sample spanning from 1952:Q3 to 2002:Q4. Row 1 confirms the results by Lettau and Ludvigson (2001) that $\widehat{cay}$ is a strong predictor of stock returns with the adjusted R^2 of 8.2%. Row 2 shows that σ_m^2 has negligible forecasting power for stock returns (row 2).[4] However, σ_m^2 becomes highly significant if $\widehat{cay}$ is also included in the forecasting equation with the adjusted R^2 of 14.7%, as shown in row 3. It should also be noted that, in the augmented model (row 3), the adjusted R^2 and the point estimates of $\widehat{cay}$ and σ_m^2 are much higher than their counterparts in rows 1 and 2. These results reflect a classic omitted-variable problem in rows 1 and 2: Although both $\widehat{cay}$ and σ_m^2 are positively related to future stock returns, they are negatively correlated with one another, as shown in table 1. Finally, row 4 shows that *rrel* provides additional information beyond $\widehat{cay}$ and σ_m^2 about future stock returns, and I find very similar results using two-period-lagged $\widehat{cay}$ in row 5.[5]

I report the estimation results using two subsample periods (1952:Q3–1977:Q4 and 1978:Q1–2002:Q4) in panels B and C, respectively. In general, the results are very similar to those reported in panel A. For example, the fore-

4. This result is sensitive to the observations of the last few years in the sample, during which σ_m^2 rose steeply, as shown in fig. 2: It becomes statistically significant if we use only the data up to 2000.

5. Adding other commonly used forecasting variables, e.g., the dividend yield, the default premium, and the term premium, does not improve the forecasting power. These results are available on request.

TABLE 2 **Forecasting One-Quarter-Ahead Excess Stock Market Returns**

	$\widehat{cay}_{t-1}$ (1)	$\widehat{cay}_{t-2}$ (2)	σ^2_{t-1} (3)	$rrel_{t-1}$ (4)	$\bar{R}^2$ (5)
			A. 1952:Q3–2002:Q4		
1	**1.911**				.082
	(4.062)				
2			2.559		.011
			(1.315)		
3	**2.639**		**5.832**		.147
	(5.278)		(3.585)		
4	**2.448**		**5.336**	**−4.365**	.163
	(5.137)		(3.372)	(−2.571)	
5		**2.065**	**5.099**	**−4.643**	.125
		(4.027)	(3.058)	(−2.579)	
			B. 1952:Q3–1977:Q4		
6	**3.105**				.163
	(4.027)				
7			**6.385**		.045
			(2.491)		
8	**3.329**		**7.687**		.234
	(4.538)		(4.245)		
9	**2.472**		**7.823**	**−8.759**	.264
	(3.080)		(4.702)	(−2.100)	
10		**1.619**	**8.172**	**−10.660**	.212
		(2.019)	(4.897)	(−3.262)	
			C. 1978:Q1–2002:Q4		
11	**1.273**				.036
	(2.352)				
12			1.207		−.006
			(.539)		
13	**2.543**		**6.066**		.096
	(3.892)		(2.984)		
14	**2.517**		**5.737**	−3.036	.100
	(3.812)		(2.849)	(−1.496)	
15		**2.324**	**5.456**	−3.309	.080
		(3.067)	(2.369)	(−1.546)	

NOTE.—We report the heteroskedasticity- and autocorrelation-adjusted t-statistics in parentheses. Regressors significant at the 5% level are in boldface.

casting ability improves substantially if I include both $\widehat{cay}$ and σ^2_m in the forecasting equation, as shown in rows 8 and 13. It should also be noted that their point estimates are strikingly similar to their full-sample counterparts in row 3, indicating a stable forecasting relation over time. This pattern explains their strong out-of-sample forecasting power presented in the next section. There are, however, some noticeable differences between the two subsamples. First, the predictability is substantially weaker in the second than in the first subsample. Second, while σ^2_m by itself is statistically significant in the first subsample (row 7), it is insignificant in the second subsample (row 12). Third, although *rrel* is statistically significant in the first subsample, it is insignificant in the second subsample. However, the two latter results are sensitive to the inclusion of observations from 2001–2 for the reasons mentioned above.

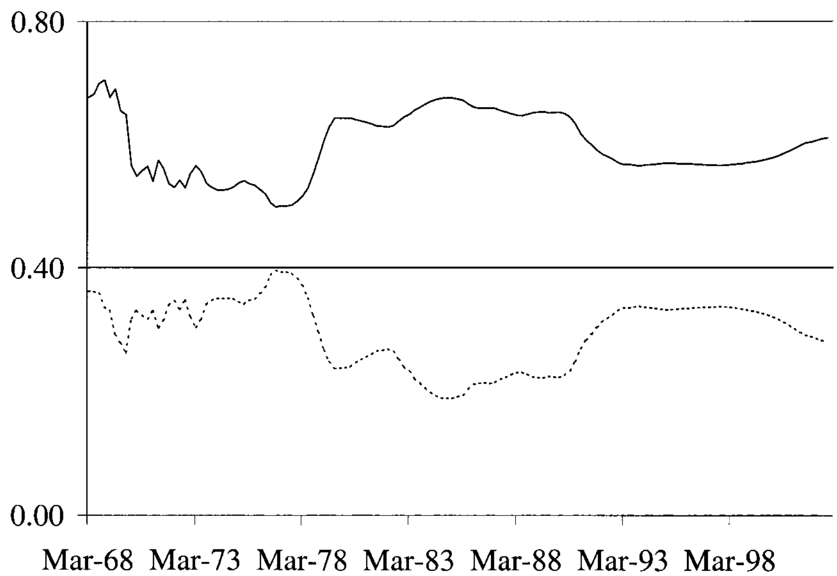

FIG. 4.—Parameters of labor income (solid line) and net worth (dashed line)

II. Out-of-Sample Forecasts

This section presents the analysis of the out-of-sample performance of various forecast models. I consider two cases. First, investors are assumed to know the cointegration parameters of *cay*, which I estimate using the full sample. They also observe consumption, labor income, and net worth without delay. This scenario is consistent with rational expectations models, in which agents have full information about the economy.[6] Second, I estimate recursively the cointegration parameters using only information available at the time of the forecast. Moreover, I lag *cay* twice, given that consumption and labor income data are available with a one-quarter delay. This scenario has appeal to practitioners, who must rely on the real-time data.[7]

Figure 4 plots the recursively estimated coefficients on labor income (solid line) and net worth (dashed line). As in Lettau and Ludvigson (2001), I estimate the cointegration parameters using dynamic least squares with eight

6. The Bureau of Economic Analysis (BEA) releases consumption and labor income data with about a one-month delay. Given that the BEA only processes but does not create data, it is possible, although unlikely, that practitioners in financial markets may obtain these data without delay. More important, *cay* is a proxy for the conditional stock market return, and practitioners may obtain similar information from alternative sources. That said, I find similar results using two-period-lagged *cay*.

7. Because consumption, net worth, and labor income data are subject to revisions, my results, which utilize the current vintage data, are potentially different from those obtained using the real-time data. While it is not clear whether the current vintage data are biased toward finding predictability, the real-time issue is beyond the scope of this paper, and I leave it for future research.

TABLE 3 **Out-of-Sample Forecast: Fixed Parameters**

	Constant (1)	$\widehat{cay}_{t-1}$ (2)	$\widehat{cay}_{t-1} + \sigma^2_{t-1}$ (3)	$\widehat{cay}_{t-1} + rrel_{t-1} + \sigma^2_{t-1}$ (4)
		A. 1968:Q2–2002:Q4		
RMSE	.0909	.0884	.0853*	.0858
MAE	.0701	.0668	.0637*	.0647
CORR	−.1838	.2332	.3135	.3379*
Sign	.5971	.6115	.6403*	.6259
Pseudo R^2		.0542	.1194*	.1091
		B. 1976:Q1–2002:Q4		
RMSE	.0843	.0852	.0829*	.0832
MAE	.0653	.0645	.0624*	.0639
CORR	−.2102	.1790	.2510*	.2336
Sign	.6296	.5926	.6389*	.6204
Pseudo R^2		−.0215	.0329*	.0259

NOTE.—This table reports five statistics for the out-of-sample test: (1) RMSE, the root mean squared error; (2) MAE, the mean of the absolute error; (3) CORR, the correlation between the forecast and the actual value; (4) Sign, the percentage of times when the forecast and the actual value have the same sign; and (5) pseudo R^2, which is equal to one minus the ratio of MSE from a forecasting model to that of the benchmark model of constant excess returns. The cointegration parameters used to calculate *cay* are estimated using the full sample. Macro variables are assumed to be available with no delay.
* The best forecast model.

leads and lags. The point estimates show large variations until the 1990s because a relatively large number of observations are required to consistently estimate the cointegration parameters. Therefore, it should not be a surprise that the forecasting ability of *cay* deteriorates significantly if the cointegration parameters are estimated recursively relative to the fixed parameters using the full sample, especially during the early period. It should also be noted that the test in the second scenario is likely to be more stringent than investors would encounter in real time, given that investors may have fairly accurate estimates of the cointegration parameters. With these caveats in mind, I report the out-of-sample forecast exercises below.

A. Fixed Cointegration Parameters

Table 3 reports the out-of-sample regression results using the fixed cointegration parameters obtained from the full sample. I analyze four forecast models, including (1) a benchmark model of constant excess returns, (2) the model using only $\widehat{cay}$, (3) the model of $\widehat{cay}$ augmented by σ^2_m, and (4) the model of $\widehat{cay}$ augmented by σ^2_m and *rrel*. Throughout the paper, I denote the model of $\widehat{cay}$ augmented by σ^2_m, which is the main focus of the analysis, as *augmented* $\widehat{cay}$. I report five commonly used forecast evaluation statistics: (1) the root mean squared error (RMSE); (2) the mean of absolute error (MAE); (3) the correlation between the forecast and the actual value (CORR); (4) the percentage of times when the forecast and the actual value have the same signs (*sign*); and (5) pseudo R^2, one minus the ratio of the mean squared error from a forecast model to the benchmark model of constant excess returns. I highlight the best forecast model for each criterion by an asterisk.

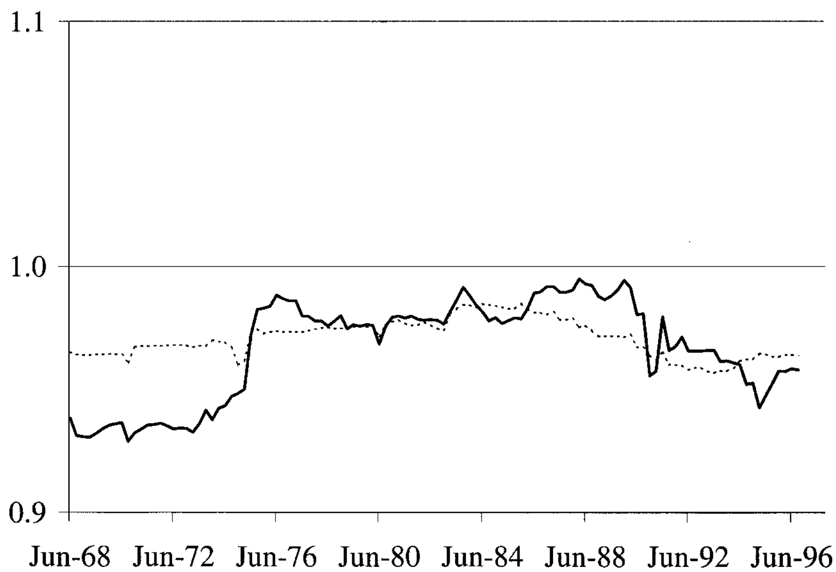

FIG. 5.—RMSE ratio of augmented *cay* to benchmark model (solid line) and to model of *cay* (dashed line): fixed parameters.

Panel A is the sample from 1968:Q2 to 2002:Q4, which is similar to the sample analyzed by Lettau and Ludvigson (2001). In the out-of-sample forecast, I first run an in-sample regression using data from 1952:Q2 to 1968:Q1 and make a forecast for 1968:Q2. Then I update the sample to 1968:Q2 and make a forecast for 1968:Q3 and so forth. Consistent with Lettau and Ludvigson, $\widehat{cay}$ (col. 2) exhibits some out-of-sample forecasting power; for example, it has a smaller RMSE than the benchmark model of constant returns (col. 1). Consistent with the in-sample regression results in table 2, its forecasting power improves dramatically by all the criteria if σ_m^2 is added to the forecasting equation (col. 3). Adding *rrel* to augmented $\widehat{cay}$ (col. 4), however, does not provide discernible improvement for the forecast performance: Overall, $\widehat{cay}$ augmented by σ_m^2 has the best out-of-sample performance.

Panel B is the subsample from 1976:Q1 to 2002:Q4. Consistent with Brennan and Xia (2002), $\widehat{cay}$ (col. 2) has a larger RMSE than the benchmark model of constant returns (col. 1) over this period. However, this result is completely reversed if I augment $\widehat{cay}$ with σ_m^2 (col. 3): Again, augmented $\widehat{cay}$ beats the other models by all criteria.

To check the robustness of the results, figure 5 plots the recursive RMSE ratio of augmented $\widehat{cay}$ (col. 3 of table 3) to the benchmark model of constant returns (col. 1; solid line) and to the model of $\widehat{cay}$ by itself (col. 2; dashed line) through time. The horizontal axis denotes the starting forecast date. For example, the value corresponding to June 1968 is the RMSE ratio over the

TABLE 4 **Out-of-Sample Forecast: Recursive Parameters**

	Constant (1)	$\widehat{cay}_{t-2}$ (2)	$\widehat{cay}_{t-2} + \sigma^2_{t-1}$ (3)	$\widehat{cay}_{t-2} + rrel_{t-1} + \sigma^2_{t-1}$ (4)
		A. 1968:Q2–2002:Q4		
RMSE	.0909	.0927	.0907	.0899*
MAE	.0701	.0704	.0678	.0674*
CORR	−.1838	.1214	.2586	.3035*
Sign	.5971	.6259*	.6187	.6187
Pseudo R^2		−.0400	.0004	.0219*
		B. 1976:Q1–2002:Q4		
RMSE	.0843*	.0879	.0862	.0853
MAE	.0653*	.0675	.0658	.0657
CORR	−.2102	.0839	.1935	.1955*
Sign	.6296*	.6204	.6204	.6111
Pseudo R^2		−.0872	−.0456	−.0239*

NOTE.—The cointegration parameters used to calculate *cay* are estimated recursively using only information available at the time of forecast. Macro variables are assumed to be available with a one-quarter delay. Also see the note of table 3.

* The best forecast model.

forecast period from 1968:Q2 to 2002:Q4. I choose the range 1968:Q2–1996:Q4 for the starting forecast date; therefore, the out-of-sample test utilizes at least 25 observations. The two ratios are always smaller than one in figure 5, indicating that (1) adding σ^2_m to the forecasting equation substantially improves the forecasting ability of $\widehat{cay}$, and (2) augmented $\widehat{cay}$ has substantial out-of-sample predictive power. In contrast, the model of $\widehat{cay}$ by itself does not always outperform the benchmark model of constant returns since the solid line is above the dashed line over various periods.

B. Recursively Estimated Cointegration Parameters

Table 4 reports the out-of-sample performance using recursively estimated $\widehat{cay}$. The exercise is the same as the case of the fixed parameters except that the cointegration parameters are estimated recursively using only information available at the time of forecast. It should be noted that consumption, labor income, and net worth are available with a one-quarter delay. For example, I first estimate the cointegration relation among consumption, net worth, and labor income and obtain the fitted $\widehat{cay}$ using data from 1952:Q2 to 1967:Q4. Then I run an in-sample forecasting regression using data from 1952:Q2 to 1968:Q1 ($\widehat{cay}$ is lagged two periods) and make a forecast for 1968:Q2. Then I update the sample to 1968:Q2 and make a forecast for 1968:Q3 and so forth. In general, the results are consistent with those in table 3. However, the forecasting ability of all models is substantially weaker in table 4 than in table 3, as expected.

In particular, for the period from 1968:Q2 to 2002:Q4, the augmented model of $\widehat{cay}$ (col. 3) performs better than the benchmark model (col. 1) and the model of $\widehat{cay}$ by itself (col. 2). Interestingly, inclusion of *rrel* (col. 4) improves

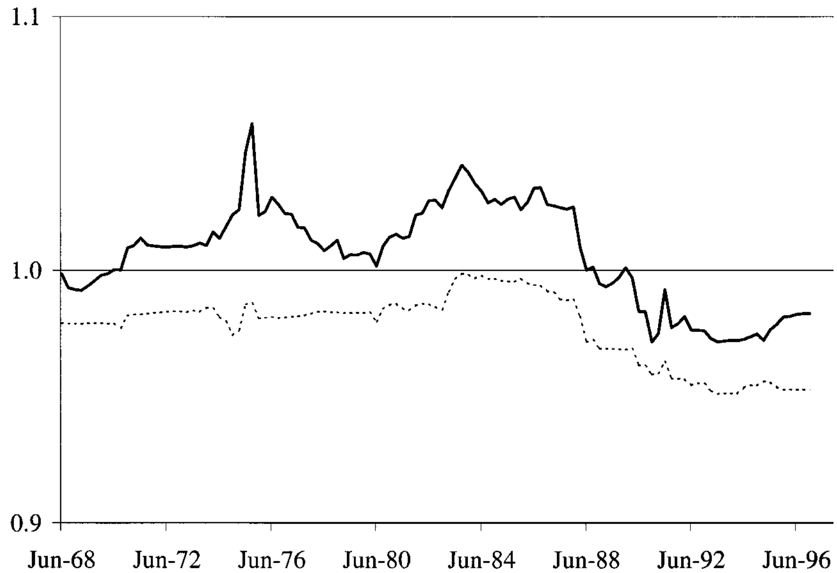

FIG. 6.—RMSE ratio of augmented *cay* to benchmark model (solid line) and to model of *cay* (dashed line): recursive parameters.

the forecasting performance of augmented $\widehat{cay}$: Overall, it has the best forecasting performance among all four models.[8] For the period 1976:Q1–2002:Q4, the benchmark model of constant returns has the smallest RMSE. Figure 6 plots the recursive RMSE ratio of augmented $\widehat{cay}$ (col. 3 of table 4) to the benchmark model of constant returns (col. 1; solid line) and to the model of $\widehat{cay}$ by itself (col. 2; dashed line) through time. The solid line remains below one after 1990, when the recursively estimated cointegration parameters become relatively stable, as shown in figure 4. Therefore, the poor out-of-sample performance of augmented $\widehat{cay}$ is mainly attributed to the large estimation errors in the cointegration parameters. Moreover, the dashed line is always below one, indicating that adding σ_m^2 to the forecasting equation substantially improves the forecasting ability of $\widehat{cay}$. It should also be noted that the solid line is always above the dashed line, indicating that the model of $\widehat{cay}$ by itself has negligible out-of-sample predictive power if the cointegration parameters are estimated recursively.

8. This result is in contrast with that in table 3, in which *rrel* provides negligible information beyond augmented *cay*. One possible explanation is that, given that recursively estimated *cay* is likely to have large measurement errors in the early period, *rrel* provides additional information in table 4 because it is closely related to "true" *cay* estimated using the full sample (as shown in table 1).

TABLE 5 One-Quarter-Ahead Forecasts of Excess Stock Market Returns: Nested Comparisons

	ENC-NEW		MSE-F	
Nested Models	Statistic (1)	Asymptotic Critical Value (2)	Statistic (3)	Asymptotic Critical Value (4)
	A. Fixed Cointegration Parameters			
1. $C+\sigma^2_{M,t-1}+cay_{t-1}$ vs. C	24.47	2.96	18.85	1.52
2. $C+\sigma^2_{M,t-1}+cay_{t-1}$ vs. $C+cay_{t-1}$	14.02	2.13	10.25	1.70
	B. Recursively Estimated Parameters			
3. $C+\sigma^2_{M,t-1}+cay_{t-2}$ vs. C	14.81	2.96	.40	1.52
4. $C+\sigma^2_{M,t-1}+cay_{t-2}$ vs. $C+cay_{t-2}$	10.22	2.13	5.97	1.70

NOTE.—This table reports two out-of-sample tests for nested forecast models: (1) the encompassing test ENC-NEW developed by Clark and McCracken (1999) and (2) the equal forecast accuracy test MSE-F developed by McCracken (1999). ENC-NEW tests the null hypothesis that the benchmark model encompasses all the relevant information for the next quarter's excess stock market return against the alternative hypothesis that past stock market variance contains additional information. MSE-F tests the null hypothesis that the benchmark model has a mean squared forecasting error that is less than or equal to the model augmented by past stock market variance against the alternative hypothesis that the augmented model has smaller mean squared forecasting error. Observations from the period 1952:Q3–1968:Q1 are used to obtain the initial in-sample estimation, and the forecasting error is calculated for the remaining period 1968:Q2–2002:Q4, recursively. For example, the forecast for 1968:Q2 is based on the estimation using the sample 1952:Q3–1968:Q1 and so forth. Columns 2 and 4 report the asymptotic 95% critical values provided by Clark and McCracken (1999). We estimate the cointegration parameters using the full sample in panel A and recursively in panel B. We compare the benchmark model of constant returns with augmented *cay* in rows 1 and 3 and compare the model of *cay* by itself with augmented *cay* in rows 2 and 4.

C. Testing Nested Forecast Models

In this subsection, I provide two formal out-of-sample tests for nested forecast models. The first is the encompassing test ENC-NEW proposed by Clark and McCracken (1999). It tests the null hypothesis that the benchmark model incorporates all the information about the next quarter's excess stock market return against the alternative hypothesis that past variance provides additional information. The second is the equal forecast accuracy test MSE-F developed by McCracken (1999). Its null hypothesis is that the benchmark model has a mean squared forecasting error less than or equal to that of the model augmented by past return; the alternative is that the augmented model has a smaller mean squared forecasting error. These two tests have also been used in Lettau and Ludvigson (2001), and Clark and McCracken (1999) find that they have the best overall power and size properties among a variety of tests proposed in the literature.

Table 5 presents the results of the out-of-sample tests. In panel A, I estimate the cointegration parameters for $\widehat{cay}$ using the full sample, and the macro variables are available without delay. I focus on two pairs of nested forecast models: the benchmark model of constant stock returns versus augmented

$\widehat{cay}$ (row 1) and the model of $\widehat{cay}$ by itself versus augmented $\widehat{cay}$ (row 2). Again, I use observations from the period 1952:Q4–1968:Q1 for the initial in-sample estimation and form the out-of-sample forecast recursively. Columns 2 and 4 report the asymptotic 95% critical value provided by Clark and McCracken (1999). I find that, in both tests, augmented $\widehat{cay}$ outperforms the model of constant returns and the model of $\widehat{cay}$ by itself at any conventional significant levels. In panel B, the cointegration parameters are estimated recursively, and the macro variables are available with a one-quarter lag. Again, I find evidence that augmented $\widehat{cay}$ outperforms the two competing models at the conventional significance level with only one exception: the MSE-F test shows that the difference between augmented $\widehat{cay}$ and the benchmark model of constant returns is not statistically significant.

III. Economic Values of Market Timing

Leitch and Tanner (1991) argue that the forecast models chosen according to statistical criteria are not necessarily the models that are profitable in timing the market. To address this issue, I investigate whether the documented predictability can be exploited to generate returns of higher mean and lower volatility than a buy-and-hold strategy offers. To conserve space, I report only the case of recursively estimated cointegration parameters, which is relevant to practitioners. Nevertheless, I find very similar results using the fixed cointegration parameters, which are available on request.

A. Switching Strategies

I adopt two widely used and relatively naive market timing strategies. The first strategy, which has been utilized by Breen and et al. (1989) and Pesaran and Timmermann (1995), among many others, requires holding stocks if the predicted excess return is positive and holding bonds otherwise. Table 6 reports the results of four trading strategies: a benchmark of buy-and-hold and three strategies based on the forecast models analyzed in tables 3 and 4. I present the mean, the standard deviation (SD), the ratio of the mean to the standard deviation (mean/SD), and the adjusted Sharpe ratio for the annualized returns on these portfolios.[9]

Over the period 1968:Q2–2002:Q4, all managed portfolios have returns of higher mean and lower standard deviation than those of the buy-and-hold strategy. For example, the managed portfolio based on the forecast model of $\widehat{cay}$ (col. 2) generates an average annual return of 13.7% with a standard deviation of 14.2%, compared with 11.3% and 18.0% respectively, for the buy-and-hold strategy (col. 1). And the adjusted Sharpe ratio of the managed portfolio is about 120% higher than the market portfolio. Therefore, even

9. As in Graham and Harvey (1997) and Johannes et al. (2002), I scale the return on the managed portfolio, e.g., through leverage, so that it has the same standard deviation as the stock market return. The scaled return is then used to calculate the Sharpe ratio in the usual way.

TABLE 6 **Switching Strategies with No Transaction Costs**

	Buy and Hold (1)	$\widehat{cay}_{t-2}$ (2)	$\widehat{cay}_{t-2} + \sigma^2_{t-1}$ (3)	$\widehat{cay}_{t-2} + rrel_{t-1} + \sigma^2_{t-1}$ (4)
			A. 1968:Q2–2002:Q4	
Mean	.1132	.1373	.1297	.1327
SD	.1801	.1421	.1645	.1560
Mean/SD	.6287	.9660	.7885	.8508
Sharpe ratio	.2873	.6247	.4471	.5094
			B. 1968:Q2–1979:Q4	
Mean	.0764	.1230	.1025	.1114
SD	.1895	.1538	.1749	.1450
Mean/SD	.4033	.7995	.5861	.7683
Sharpe ratio	.0840	.4802	.2669	.4490
			C. 1980:Q1–1989:Q4	
Mean	.1700	.1879	.1879	.1850
SD	.1781	.1634	.1634	.1692
Mean/SD	.9543	1.1468	1.1468	1.0929
Sharpe ratio	.4795	.6718	.6719	.6181
			D. 1990:Q1–2002:Q4	
Mean	.1029	.1114	.1096	.1117
SD	.1737	.1099	.1556	.1556
Mean/SD	.5924	1.0134	.7046	.7181
Sharpe ratio	.3354	.7564	.4477	.4611

NOTE.—The table reports returns on switching strategies, which require holding stocks if the predicted excess return is positive and holding bonds otherwise. I present four statistics for the annualized return on the managed portfolio, including the mean, the standard deviation (SD), the ratio of the mean to the standard deviation (mean/SD), and the adjusted Sharpe ratio. As in Graham and Harvey (1997), I scale the return on the managed portfolio, e.g., through leverage, so that it has the same standard deviation as the market return. The scaled return is then used to calculate the Sharpe ratio in the usual way. The cointegration parameters used to calculate *cay* are estimated recursively using only information available at the time of forecast. Macro variables are assumed to be available with a one-quarter delay.

though the out-of-sample forecasting ability of $\widehat{cay}$ is statistically negligible as shown in table 4, it is economically important. My results thus confirm Leitch and Tanner's (1991) skepticism of using statistical criteria such as RMSE for forecast evaluation. Also, in contrast with the results of table 4, the model augmented with σ^2_m and *rrel* (col. 4) has an adjusted Sharpe ratio lower than the model that uses $\widehat{cay}$ only. This is also true for the model augmented with σ^2_m (col. 3). As I show below, these results reflect the fact that information is not used efficiently in the switching strategy.

I find very similar patterns in the three subsample periods, which are reported in panels B–D of table 6. However, the performance of the managed portfolio relative to the benchmark fluctuates widely over time, which is consistent with the finding of Pesaran and Timmermann (1995). For example, for the market timing strategy based on $\widehat{cay}$ only, one observes the biggest improvement in the 1970s: The managed portfolio has an adjusted Sharpe ratio of 0.48, compared with 0.08 for the market portfolio. In contrast, the managed portfolio has an adjusted Sharpe ratio of 0.67 (0.76) for the period

1980:Q1–1989:Q4 (1990:Q1–2002:Q4), compared with 0.48 (0.34) for the market portfolio. I find a similar pattern for the other forecast models.

Figure 7 provides some details of the strategy based on augmented $\widehat{cay}$ (col. 3 of table 6). The upper panel plots the weight of stocks in the managed portfolio, which assumes two values of zero (100% of bonds) and one (100% of stocks). Interestingly, investors did not have to rebalance the portfolio very often, especially during the stock market run-up in the 1980s and 1990s. The lower panel shows that, by using our forecasting variables to time the market, investors avoid some large downward movements in the stock market, for example, around the 1973 oil shock. Finally, the middle panel plots the value of a \$100 investment in a market index (dashed line) and in the managed portfolio (solid line), respectively, starting from 1968:Q2. The latter is always higher than the former. By the end of 2002:Q4, the managed portfolio is worth \$5,338, compared with \$2,793 for the buy-and-hold strategy.

Table 7 investigates the effect of a proportional transaction cost of 25 basis points. For example, when investors switch from stocks to bonds or vice versa, they have to pay a fee of 0.25% of the value of their portfolios. It should be noted that a 25-basis-point fee is in the upper range of transaction costs for the market index (e.g., Balduzzi and Lynch 1999). In a comparison with the results in table 6, I find that transaction costs have a small impact on the performance of the managed portfolio. This result should not be a surprise because investors did not rebalance the managed portfolio very often, as shown in figure 7.

B. Choosing Optimal Portfolio Weights

In the second strategy, which has been adopted by Johannes et al. (2002), among others, I allocate wealth among stocks and bonds using the static CAPM. Specifically, I invest a fraction of total wealth,

$$\omega_t = \frac{1}{\gamma}\frac{E_t[R_{t+1} - R_f]}{E_t\sigma^2_{m,t+1}},$$

in stocks and a fraction $1 - \omega_t$ in bonds, where γ is a measure of the investor's relative risk aversion, $E_t[R_{t+1} - R_f]$ is the predicted value from the excess return forecasting regression, and $E_t\sigma^2_{m,t+1}$ is the conditional variance measured by the fitted value from a regression of realized variance, $\sigma^2_{m,t+1}$, on a constant and its two lags. Compared with the first strategy, this strategy is plausible because it incorporates the information of not only signs but also the magnitude of the predicted excess return normalized by its variance. For simplicity, I ignore the estimation uncertainty, on which Johannes et al. offer some detailed discussion. I also assume that ω_t is in the range [0, 1] or that investors are not allowed to short sell stocks or borrow from bond markets because those transactions might be infeasible in practice owing to high costs. It should be noted that the profitability of timing strategies should in principle be lower

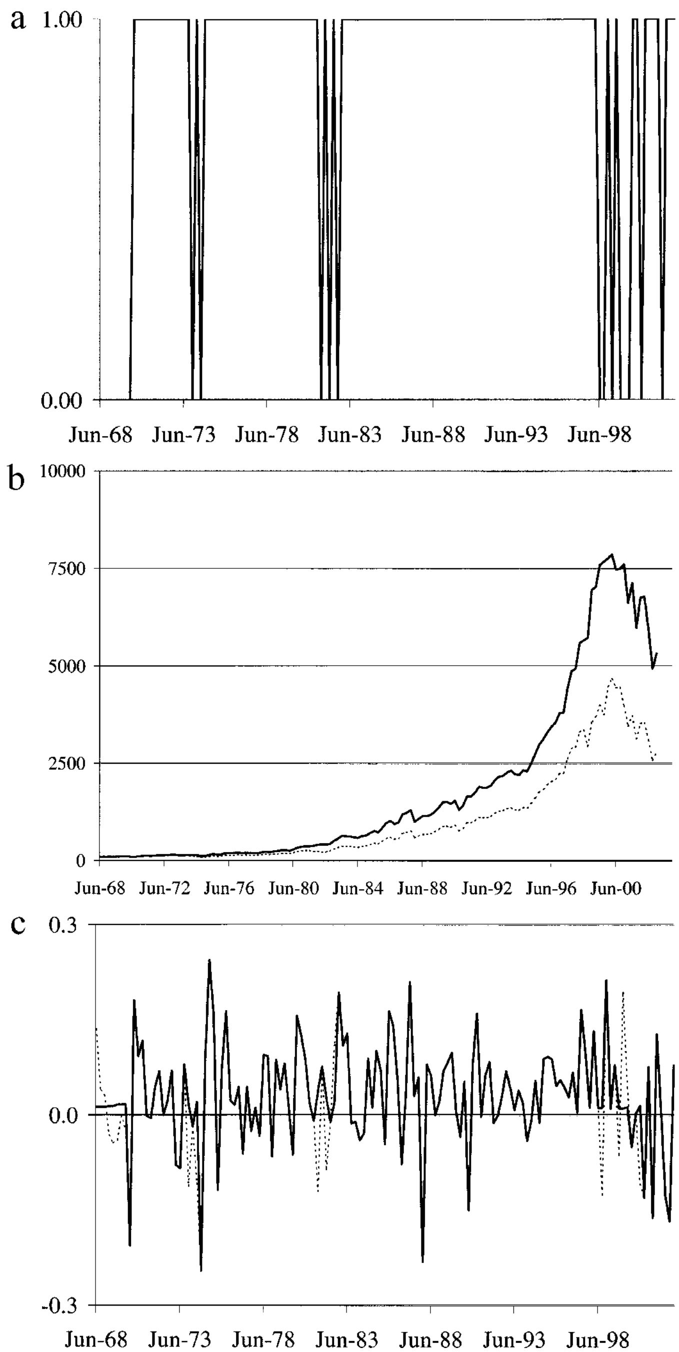

FIG. 7.—Switching strategies. *a*, Weight of stocks in managed portfolio. *b*, Values of managed portfolio (solid line) vs. market portfolio (dashed line). *c*, Returns on managed portfolio (solid line) vs. market portfolio (dashed line).

TABLE 7 Switching Strategies with Transaction Costs

	Buy and Hold (1)	$\widehat{cay}_{t-2}$ (2)	$\widehat{cay}_{t-2} + \sigma^2_{t-1}$ (3)	$\widehat{cay}_{t-2} + rrel_{t-1} + \sigma^2_{t-1}$ (4)
	A. 1968:Q2–2002:Q4			
Mean	.1132	.1363	.1282	.1315
SD	.1801	.1422	.1647	.1562
Mean/SD	.6287	.9585	.6917	.7031
Sharpe ratio	.2873	.6171	.4369	.5005
	B. 1968:Q2–1979:Q4			
Mean	.0764	.1224	.1015	.1100
SD	.1895	.1540	.1754	.1454
Mean/SD	.4033	.7944	.5787	.7564
Sharpe ratio	.0840	.4752	.2595	.4372
	C. 1980:Q1–1989:Q4			
Mean	.1700	.1862	.1862	.1840
SD	.1781	.1636	.1636	.1694
Mean/SD	.9543	1.1380	1.1380	1.0856
Sharpe ratio	.4795	.6632	.6632	.6108
	D. 1990:Q1–2002:Q4			
Mean	.1029	.1104	.1077	.1105
SD	.1737	.1101	.1558	.1556
Mean/SD	.5924	1.0031	.6913	.7104
Sharpe ratio	.3354	.7462	.4344	.4534

NOTE.—I assume that investors have to pay a proportional transaction cost of 25 basis points when they switch from stocks to bonds or vice versa. The other specifications are the same as in table 6.

under these assumptions than otherwise because they reduce the set of investment opportunities and lead to a lower mean-variance frontier.

Table 8 reports the statistics for returns on the managed portfolio based on various forecast models. In the calculation of the optimal weight for stocks, I assume that $\gamma = 5$.[10] As expected, the portfolio based on augmented $\widehat{cay}$ (col. 3) has substantially higher Sharpe ratios than those reported in table 6 for the switching strategy. For example, over the period 1968:Q2–2002:Q4, the Sharpe ratio is 0.59 if investors choose portfolio weight optimally, compared with 0.45 for the switching strategy. Nevertheless, the other results are very similar to those reported in table 6. For example, market timing strategies based on models using $\widehat{cay}$ as a forecasting variable generate returns of higher mean and lower volatility than the buy-and-hold strategy. Also, the relative performance of market timing strategies fluctuates widely over time and is the most effective in the 1970s.

Figure 8 provides some details of the market timing strategy based on augmented $\widehat{cay}$ (col. 3 of table 7). Again, the upper panel plots the weight of stocks in the managed portfolio, which is very similar to that of figure 7 except that the weight occasionally takes a value between zero and one. The

10. The results are not sensitive to reasonable variations in γ.

TABLE 8. Choosing Optimal Portfolio Weights with No Transaction Costs

	Buy and Hold (1)	$\widehat{cay}_{t-2}$ (2)	$\widehat{cay}_{t-2} + \sigma^2_{t-1}$ (3)	$\widehat{cay}_{t-2} + rrel_{t-1} + \sigma^2_{t-1}$ (4)
		A. 1968:Q2–2002:Q4		
Mean	.1132	.1156	.1355	.1381
SD	.1801	.1235	.1454	.1459
Mean/SD	.6287	.9367	.9316	.9467
Sharpe ratio	.2873	.5953	.5902	.6053
		B. 1968:Q2–1979:Q4		
Mean	.0764	.0941	.1211	.1353
SD	.1895	.1243	.1454	.1242
Mean/SD	.4033	.7570	.8332	1.0887
Sharpe ratio	.0840	.4378	.5140	.7695
		C. 1980:Q1–1989:Q4		
Mean	.1700	.1476	.1690	.1741
SD	.1781	.1446	.1550	.1653
Mean/SD	.9543	1.0207	1.0906	1.0532
Sharpe ratio	.4795	.5459	.6158	.5784
		D. 1990:Q1–2002:Q4		
Mean	.1029	.1106	.1227	.1129
SD	.1737	.1049	.1396	.1494
Mean/SD	.5924	1.0535	.8786	.7558
Sharpe ratio	.3354	.7965	.6217	.4988

NOTE.—The table reports returns on strategies for choosing optimal portfolio weights. In particular, investors allocate a fraction of total wealth,

$$\omega_t = \frac{1}{\gamma}\frac{E_t[R_{t+1} - R_f]}{E_t\sigma^2_{m,t+1}},$$

in stocks and a fraction $1 - \omega_t$ in bonds, where γ is a measure of the investor's relative risk aversion, $E_t[R_{t+1} - R_f]$ is the predicted value from the excess return forecasting regression, and $E_t\sigma^2_{m,t+1}$ is the conditional variance measured by the fitted value from a regression of $\sigma^2_{m,t+1}$ on a constant and its two lags. For simplicity, I ignore the estimation uncertainty and assume that ω_t is in the range [0, 1]. The cointegration parameters used to calculate *cay* are estimated recursively using only information available at the time of forecast. Macro variables are assumed to be available with a one-quarter delay. Also see the note of table 6.

lower panel plots the return on the managed portfolio (solid line) as well as the market return (dashed line). Compared with the first strategy plotted in figure 7, the second strategy successfully avoids additional major downward movements in the stock market. The middle panel shows that a $100 initial investment in the managed portfolio grows to $7,227 by the end of year 2002, which is over 2.5 times as much as the market portfolio. Again, table 9 shows that transaction costs have small effects on the performance of the managed portfolio.

C. Some Further Tests

Cumby and Modest (1987) propose a formal test of market timing ability by regressing the realized excess return, $r_{m,t+1} - r_{f,t+1}$, on a constant and an in-

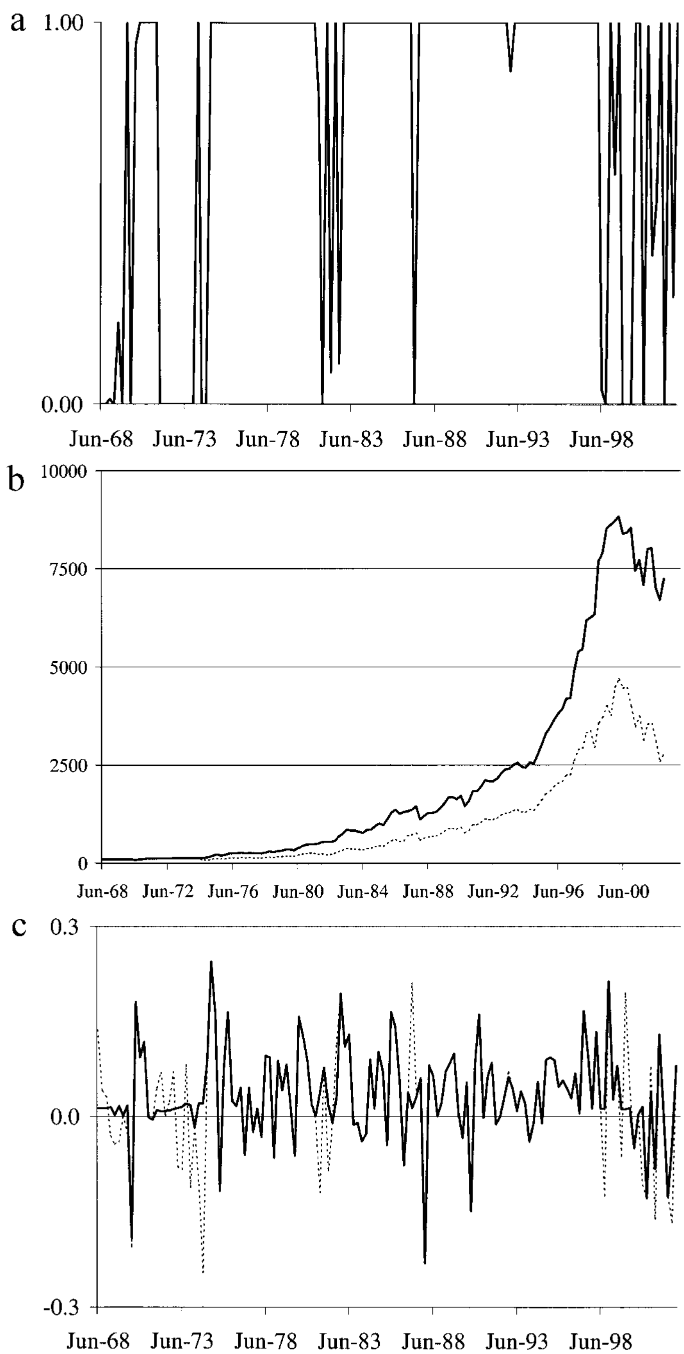

FIG. 8.—Choosing optimal portfolio weights. *a*, Weight of stocks in managed portfolio. *b*, Values of managed portfolio (solid line) vs. market portfolio (dashed line). *c*, Returns on managed portfolio (solid line) vs. market portfolio (dashed line).

TABLE 9. Choosing Optimal Portfolio Weights with Transaction Costs

	Buy and Hold (1)	$\widehat{cay}_{t-2}$ (2)	$\widehat{cay}_{t-2} + \sigma^2_{t-1}$ (3)	$\widehat{cay}_{t-2} + rrel_{t-1} + \sigma^2_{t-1}$ (4)
	A. 1968:Q2–2002:Q4			
Mean	.1132	.1141	.1335	.1363
SD	.1801	.1234	.1456	.1459
Mean/SD	.6287	.9245	.9171	.9343
Sharpe ratio	.2873	.5831	.5757	.5929
	B. 1968:Q2–1979:Q4			
Mean	.0764	.0921	.1195	.1330
SD	.1895	.1240	.1457	.1239
Mean/SD	.4033	.7430	.8204	1.0729
Sharpe ratio	.0840	.4237	.5012	.7537
	C. 1980:Q1–1989:Q4			
Mean	.1700	.1455	.1670	.1726
SD	.1781	.1447	.1549	.1656
Mean/SD	.9543	1.0056	1.0784	1.0424
Sharpe ratio	.4795	.5308	.6036	.5676
	D. 1990:Q1–2002:Q4			
Mean	.1029	.1099	.1204	.1115
SD	.1737	.1052	.1399	.1496
Mean/SD	.5924	1.0452	.8609	.7451
Sharpe ratio	.3354	.7882	.6040	.4881

NOTE.—I assume that investors have to pay a proportional transaction cost, which is equal to 0.25% times the absolute value of the change in the weight of stocks in the managed portfolio. The other specifications are the same as in table 8. Also see the note in table 6.

dicator variable, I_t, which is equal to one if $r_{m,t+1} - r_{f,t+1}$ is expected to be positive and is equal to zero otherwise, as in the following equation:

$$r_{m,t+1} - r_{f,t+1} = a + b \times I_t + \varepsilon_{t+1}. \tag{1}$$

Under the null hypothesis of no market timing ability, the coefficient of the indicator variable, b, should not be statistically different from zero. Table 10 reports the regression results. Over the period 1968:Q2–2002:Q4, the null hypothesis of no market timing ability is rejected for all the forecast models.

I also investigate whether the CAPM and the Fama-French model can explain returns on the managed portfolio. For the CAPM, I run regressions of excess returns on the managed portfolio, $r_{mp,t+1} - r_{f,t+1}$, on a constant and a single factor of excess stock market returns, as in equation (2). I include two additional factors: the return on a portfolio that is long in small stocks and short in large stocks (SMB) and the return on a portfolio that is long in high book-to-market stocks and short in low book-to-market stocks (HML) for the Fama-French model:[11]

$$r_{mp,t+1} - r_{f,t+1} = \alpha + \sum \beta_i f_i + \varepsilon_{t+1}. \tag{2}$$

11. SMB and HML are obtained from Kenneth French at Dartmouth College.

TABLE 10 Cumby and Modest (1987) Market Timing Ability Test: 1968:Q2–2002:Q4

	$\widehat{cay}_{t-2}$ (1)	$\widehat{cay}_{t-2} + \sigma^2_{t-1}$ (2)	$\widehat{cay}_{t-2} + rrel_{t-1} + \sigma^2_{t-1}$(3)
a	−.025	−.027	−.022
	(−1.352)	(−1.433)	(−1.364)
b	**.050**	**.047**	**.448**
	(2.490)	(2.271)	(2.453)

NOTE.—The table reports the Cumby and Modest (1987) market timing ability test, eq. (1), where $r_{m,t+1} - r_{f,t+1}$ is the realized excess stock market return, and I_t is an indicator variable that is equal to one if $r_{m,t+1} - r_{f,t+1}$ is expected to be positive and is equal to zero otherwise. Regressors significant at the 5% level are in boldface. The cointegration parameters used to calculate *cay* are estimated recursively using only information available at the time of forecast. Macro variables are assumed to be available with a one-quarter delay.

Under the joint null hypothesis that (1) the CAPM or the Fama-French model is the correct model and (2) the managed portfolio is rationally priced, the constant term, α, should not be statistically different from zero. I report the regression results in table 11. Panels A and B are the cases of no transaction costs. For both strategies, the CAPM cannot explain returns on the managed portfolio over the period 1968:Q2–2002:Q4. The Fama-French model explains the returns somewhat better; however, α is still significant for $\widehat{cay}$ augmented by σ^2_m and *rrel* (col. 3), is marginally significant for $\widehat{cay}$ augmented by σ^2_m(col. 2) in panel B, and is marginally significant for $\widehat{cay}$ by itself (col. 1) in panel A. Again, I find essentially the same results if I incorporate a proportional transaction cost of 25 basis points in panels C and D.

Finally, I calculate the certainty equivalence gain of holding the managed portfolio, as in Fleming, Kirby, and Ostdiek (2001). I assume that the utility function has the form

$$U = W_0\left[\sum_{t=0}^{T-1} R_{p,t+1} - \frac{\gamma}{2(1+\gamma)} R^2_{p,t+1}\right], \tag{3}$$

where W_0 is initial wealth and $R_{p,t+1}$ is the return on the agent's portfolio. The certainty equivalence gain, Δ, is defined in equation (4) as the fee that an investor would pay in exchange for holding the managed portfolio that pays a rate of return $R_{mp,t+1}$; otherwise, he holds the market portfolio that pays $R_{m,t+1}$:

$$\sum_{t=0}^{T-1} (R_{mp,t+1} - \Delta) - \frac{\gamma}{2(1+\gamma)} (R_{mp,t+1} - \Delta)^2 = \sum_{t=0}^{T-1} R_{m,t+1} - \frac{\gamma}{2(1+\gamma)} R^2_{m,t+1}. \tag{4}$$

Table 12 shows that the certainty equivalent gain of holding the managed portfolio is quite substantial, usually ranging from 2% to 3%. Moreover, transaction costs have a small effect on the results.

TABLE 11 **Jensen's α Tests: 1968:Q2–2002:Q4**

	$\widehat{cay}_{t-2}$	$\widehat{cay}_{t-2} + \sigma^2_{t-1}$	$\widehat{cay}_{t-2} + rrel_{t-1} + \sigma^2_{t-1}$
	A. Switching Strategies with No Transaction Costs		
CAPM	**.011**	**.006**	**.008**
	(2.791)	(2.154)	(2.245)
Fama-French	.008	.005	.006
	(1.955)	(1.434)	(1.606)
	B. Optimal Weight Strategies with No Transaction Costs		
CAPM	.007	**.010**	**.010**
	(1.895)	(2.678)	(2.768)
Fama-French	.004	.008	**.008**
	(1.081)	(1.926)	(2.094)
	C. Switching Strategies with Transaction Costs		
CAPM	**.011**	**.006**	**.008**
	(2.725)	(2.033)	(2.162)
Fama-French	.008	.004	.006
	(1.894)	(1.328)	(1.530)
	D. Optimal Weight Strategies with Transaction Costs		
CAPM	.007	**.009**	**.010**
	(1.792)	(2.542)	(2.648)
Fama-French	.004	.007	**.008**
	(.994)	(1.801)	(1.984)

NOTE.—The table reports Jensen's α test for returns on the managed portfolio. As in eq. (2), I run a regression of excess return on the managed portfolio, $r_{mp,t+1} - r_{f,t+1}$, on a constant and risk factors. The risk factor includes only the excess stock market return in the CAPM. For the Fama-French model, I include two additional factors: The return on a portfolio that is long in small stocks and short in large stocks and the return on a portfolio that is long in high book-to-market stocks and short in low book-to-market stocks. Regressors significant at the 5% level are in boldface. The cointegration parameters used to calculate *cay* are estimated recursively using only information available at the time of forecast. Macro variables are assumed to be available with a one-quarter delay.

IV. Conclusion

In this paper, I show that the out-of-sample predictability of stock market returns is both statistically and economically significant. More important, in sharp contrast to early empirical work, I find that, in conjunction with the consumption-wealth ratio, stock market volatility has strong forecasting power for stock market returns—a key implication of the CAPM. My results thus suggest that stock return predictability is not inconsistent with rational pricing, a point that has been emphasized by Campbell and Cochrane (1999) and Guo (2004), among others.

I also want to stress that the forecasting ability of the consumption-wealth ratio is well motivated: It reflects a liquidity premium due to limited stock market participation, as in Guo (2004). In particular, it helps explain why the early authors failed to find significant forecasting power of volatility for stock returns: The risk and liquidity premiums are negatively related in the post–World War II sample. It also sheds light on the puzzling negative risk-return relation documented in the early literature: Guo (2002b) shows that

TABLE 12 Certainty Equivalence Gains from Holding Managed Portfolio: 1968:Q2–2002:Q4

$\widehat{cay}_{t-2}$ (1)	$\widehat{cay}_{t-2} + \sigma^2_{t-1}$ (2)	$\widehat{cay}_{t-2} + rrel_{t-1} + \sigma^2_{t-1}$ (3)
A. Switching Strategies with No Transaction Costs		
.0292	.0188	.0229
B. Optimal Weight Strategies with No Transaction Costs		
.0097	.0270	.0296
C. Switching Strategies with Transaction Costs		
.0282	.0172	.0216
D. Optimal Weight Strategies with Transaction Costs		
.0082	.0251	.0278

NOTE.—I assume that the utility function has the form of eq. (3), where W_0 is initial wealth and $R_{p,t+1}$ is the return on the agent's portfolio. The certainty equivalence gain, Δ, is defined in eq. (4), which is the fee that an investor would pay in exchange for holding the managed portfolio that pays a rate of return $R_{mp,t+1}$; otherwise, he holds the market portfolio that pays $R_{m,t+1}$, as in eq. (4). The cointegration parameters used to calculate *cay* are estimated recursively using only information available at the time of forecast. Macro variables are assumed to be available with a one-quarter delay.

market risk is indeed positively priced if one controls for the liquidity premium.

It is important to notice that evidence that the CAPM and the Fama-French model cannot explain the return on the managed portfolio does not necessarily pose a challenge to rational asset pricing theories. The reason is that, as shown by Merton (1973) and Campbell (1993), among others, a hedge for investment opportunity changes is also an important determinant of expected asset return, in addition to market risk. Using the same forecasting variables as in this paper, Guo (2002a) shows that Campbell's (1993) intertemporal CAPM is quite successful in explaining the cross section of stock returns, including the momentum profit, which also challenges the CAPM and the Fama-French model.[12]

Overall, stock return predictability documented in this paper has important implications for asset pricing and portfolio management and warrants attention in future research.

Appendix

cay versus *tay*

Brennan and Xia (2002) suggest that the predictive power of *cay* is spurious because, if calendar time is used in place of consumption, the resulting cointegration error, *tay*—an inanimate variable—performs as well as or better than *cay*. In their reply,

12. Although the Fama-French model is intended to capture the hedge for investment opportunity changes, its choice of additional risk factors is admittedly ad hoc.

TABLE A1 Forecasting One-Quarter-Ahead Excess Stock Market Returns: *cay* versus *tay*

	$\widehat{tay}_{t-1}$ (1)	$\widehat{tay}_{t-2}$ (2)	$\widehat{cay}_{t-1}$ (3)	$\widehat{cay}_{t-2}$ (4)	σ^2_{t-1} (5)	$rrel_{t-1}$ (6)	$\bar{R}^2$ (7)
1	.002		1.190				.089
	(1.749)		(1.801)				
2	.001		**2.152**		**5.607**		.147
	(.999)		(2.969)		(3.439)		
3	.001		**2.238**		**5.262**	**−4.116**	.160
	(.457)		(3.045)		(3.312)	(−2.418)	
4		.001		**1.735**	**5.033**	**−4.258**	.123
		(.797)		(2.367)	(2.990)	(−2.402)	

NOTE.—I report the heteroskedasticity- and autocorrelation-adjusted *t*-statistics in parentheses. Regressors significant at the 5% level are in boldface.

Lettau and Ludvigson (2002) argue that, given that 99% of variations of consumption are explained by a time trend, *tay*, a seemingly inanimate variable, has more economic content than it appears. However, I still need to show that *cay* performs at least as well as *tay*, which I discuss in this appendix.

Table A1 presents the in-sample regression results using the full sample from 1952: Q3 to 2002:Q4. Consistent with Brennan and Xia, row 1 shows that $\widehat{cay}$ becomes statistically insignificant at the 5% level if *tay* is also included in the forecasting equation. However, this result is dramatically reversed if one adds σ^2_m to the forecasting equation: $\widehat{cay}$ drives out *tay* in row 2. I find the same results if I also add *rrel* to the forecasting equation (row 3) or use two-period-lagged $\widehat{cay}$ and *tay* (row 4).

I also repeat the exercises of Sections II and III using *tay* in place of *cay*. Consistent with the in-sample regression results, I find that *cay* always outperforms *tay* in the out-of-sample tests if augmented with σ^2_m. To conserve space, these results are not reported here but are available on request. Therefore, although the results by Brennan and Xia are interesting because they reflect an unstable relation between *cay* and excess stock market returns as a result of the omitted-variable problem documented in this paper, they do not pose a challenge to the forecasting power of *cay*.

References

Ang, A., and G. Bekaert. 2001. Stock return predictability: Is it there? Unpublished working paper, Columbia University, Graduate School of Business.

Balduzzi, P., and A. Lynch. 1999. Transaction costs and predictability: Some utility cost calculations. *Journal of Financial Economics* 52:47–78.

Bernanke, B., and M. Gertler. 1989. Agency costs, net worth, and business fluctuations. *American Economic Review* 79:14–31.

Bossaerts, P., and P. Hillion. 1999. Implementing statistical criteria to select return forecasting models: What do we learn? *Review of Financial Studies* 12:405–28.

Breen, W., L. Glosten, and R. Jagannathan. 1989. Economic significance of predictable variations in stock index returns. *Journal of Finance* 44:1177–89.

Brennan, M., and Y. Xia. 2002. *tay*'s as good as *cay*. Unpublished working paper, University of California, Los Angeles, Anderson School of Business.

Campbell, J. 1987. Stock returns and the term structure. *Journal of Financial Economics* 18: 373–99.

———. 1993. Intertemporal asset pricing without consumption data. *American Economic Review* 83:487–512.

Campbell, J., and J. Cochrane. 1999. By force of habit: A consumption-based explanation of aggregate stock market behavior. *Journal of Political Economy* 107:205–51.

Campbell, J., M. Lettau, B. Malkiel, and Y. Xu. 2001. Have individual stocks become more volatile? An empirical exploration of idiosyncratic risk. *Journal of Finance* 56:1–43.

Campbell, J., A. Lo, and C. MacKinlay. 1997. *The econometrics of financial markets*. Princeton, NJ: Princeton University Press.

Clark, T., and M. McCracken. 1999. Tests of equal forecast accuracy and encompassing for nested models. Unpublished working paper, Federal Reserve Bank of Kansas City.

Cumby, E., and D. Modest. 1987. Testing for market timing ability: A framework for evaluation. *Journal of Financial Economics* 19:169–89.

Fama, E., and K. French. 1989. Business conditions and expected returns on stocks and bonds. *Journal of Financial Economics* 25:23–49.

———. 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33:3–56.

Fleming, J., C. Kirby, and B. Ostdiek. 2001. The economic value of volatility timing. *Journal of Finance* 56:329–52.

Goyal, A., and I. Welch. 2003. Predicting the equity premium with dividend ratios. *Management Science* 49:639–54.

Graham, J., and C. Harvey. 1997. Grading the performance of market-timing newsletters. *Financial Analysts Journal* 53:54–66.

Guo, H. 2002a. Time-varying risk premia and the cross section of stock returns. Working Paper no. 2002-013A, Federal Reserve Bank of St. Louis.

———. 2002b. Understanding the risk-return tradeoff in the stock market. Working Paper no. 2002-001A, Federal Reserve Bank of St. Louis.

———. 2004. Limited stock market participation and asset prices in a dynamic economy. *Journal of Financial and Quantitative Analysis* 39:495–516.

Inoue, A., and L. Kilian. 2002. In-sample or out-of-sample tests of predictability: Which one should we use? Unpublished working paper, University of Michigan, Department of Economics.

Johannes, M., N. Polson, and J. Stroud. 2002. Sequential optimal portfolio performance: Market and volatility timing. Unpublished working paper, University of Chicago, Graduate School of Business.

Leitch, G., and E. Tanner. 1991. Economic forecast evaluation: Profits versus conventional error measures. *American Economic Review* 81:580–90.

Lettau, M., and S. Ludvigson. 2001. Consumption, aggregate wealth, and expected stock returns. *Journal of Finance* 56, no. 3:815–49.

———. 2002. *tay*'s as good as *cay*: Reply. Unpublished working paper, New York University, Department of Economics.

McCracken, M. 1999. Asymptotics for out-of-sample tests of causality. Unpublished working paper, Louisiana State University.

Merton, R. 1973. An intertemporal capital asset pricing model. *Econometrica* 41:867–87.

———. 1980. On estimating the expected return on the market: An exploratory investigation. *Journal of Financial Economics* 8:323–61.

Patelis, A. 1997. Stock return predictability and the role of monetary policy. *Journal of Finance* 52:1951–72.

Pesaran, M., and A. Timmermann. 1995. Predictability of stock returns: Robustness and economic significance. *Journal of Finance* 50:1201–28.

Schwert, G. 1990. Indexes of stock prices from 1802 to 1987. *Journal of Business* 63:399–426.