Market Timing with *Cay*

Using deviation from the long-run aggregate log consumption-wealth ratio.

Sandro C. Andrade, Ilona Babenko, and Yuri Tserlukevich

Sandro C. Andrade, Ilona Babenko, and Yuri Tserlukevich are at the Haas School of Business of the University of California at Berkeley
andrade@haas.berkeley.edu

The debate about market return predictability and whether it can be successfully used in market timing strategies has attracted academicians and practitioners for a long time. We use a relatively new forecasting variable to create timing strategies.

The variable *cay* is the deviation from the long-run aggregate log consumption-wealth ratio of the U.S. economy, introduced by Lettau and Ludvigson [2001a]. *Cay* needs to be estimated from a time series of data because human capital, a part of total wealth, is not observable, and also because the long-run equilibrium consumption-wealth ratio is unknown. *Cay* is empirically constructed as the residual term from an estimate of the cointegration relation of aggregate consumption, total asset wealth, and aggregate labor income.[1]

Campbell [1996] has argued that a variable that does well in cross-section will not likely do well in time series. *Cay* generates interest in academic circles because it works well in both time series and cross-sectionally.

Lettau and Ludvigson [2001a] show that *cay* forecasts excess returns of the equity market better than other variables such as the aggregate dividend-price ratio and short-term interest rates. Lettau and Ludvigson [2001b] show that *cay* helps to explain the variation of average returns across stock portfolios when it is used as a conditioning variable in a theoretical asset pricing model, the consumption CAPM.[2]

Cay's good performance can be justified by an intuitive argument based on the consumption smoothing behavior of individuals. In expectation of higher future returns, investors consume more today than might be expected, given their wealth level. The consumption-wealth ratio would aggregate and summarize expectations about the effect of future returns on total wealth.[3]

We go beyond the Lettau and Ludvigson finding, and ask whether the predictive power of *cay* can be translated into superior investment performance. Lettau and Ludvigson [2001a] find that *cay* explains about 9% of excess stock return variability, as measured by the R-squared. They do not, however, assess to what extent this explanatory power can be useful in generating abnormal returns in market timing strategies.

We construct market timing strategies using *cay* as a predictive variable, and evaluate their performance using several common tests. We compare two sets of results. First, we assume that the market timer does not need to estimate the cointegration relationship that yields *cay*; that is, the timer has access to the current *cay* at every time. Second, we assume that the market timer has to estimate the current *cay* using only data available at the moment, as would be the case in real-world money management.

Our results indicate that, despite some potential for successful market timing strategies based on the information embedded in *cay*, practical implementation remains debatable.

METHODOLOGY

We investigate whether a hypothetical mean-variance mutual fund manager is able to time the market using predictability regressions with quarterly S&P 500 excess returns. At the end of each quarter, the manager makes an out-of-sample forecast of the expected excess return of the index for the next quarter. The manager uses this forecast to build an optimal portfolio composed of two assets: the S&P 500 index and a three-month T-bill.[4]

The main forecasting variable is *cay*, which is constructed as the cointegrating residual for consumption c, asset wealth a, and labor income y. In our first set of results, which we call *full-sample cay lagged once*, the cointegration relation is estimated using the full sample from 1951:4 through 2003:2. We follow Lettau and Ludvigson [2001a] and use dynamic least squares with eight leads and lags, which requires the estimation of 37 parameters.

This yields the formula for the estimation of *cay*:

$$cay_t = c_t - 0.2711a_t - 0.6185y_t - 0.7311 \quad (1)$$

In the first set of results, we assume the manager knows this formula as well as the end-of-current-quarter values of c, a, and y, and therefore is able to compute the end-of-current-quarter value of *cay*. The time series of *cay* will be used to forecast the next-period excess return of the S&P 500.

In our second set of results, *called recursive cay lagged twice*, the manager has to find the coefficients of the formula by estimating the cointegration relation at each time, using only currently available information. Since macroeconomic data for the current quarter are not released until later in the next quarter, we also assume the manager does not have the end-of-current-quarter values of c, a, and y. Therefore, the latest estimate of *cay* will be the one for the end of the previous quarter.

To guarantee at least 20 years of data for the recursive estimation of *cay*, the initial portfolios are built at the end of the first quarter of 1972; the last rebalancing occurs at the end of the first quarter of 2003. We denote this recursively estimated *cay* by $\widehat{cay}$.

Since *cay*'s effectiveness depends mostly on long-run dynamics, we also add the slope of the interest rate term structure (SLOPE) and the detrended short-term interest rate (RREL) to the predictive regressions. These variables are meant to pick up shorter-run business cycle effects. SLOPE is the yield spread of a ten-year Treasury over a three-month T-bill. Following Campbell [1987], we define RREL as the current one-month T-bill yield minus its 12-month moving average.

The manager considers four strategies. Strategy A uses *cay* only. Strategy B uses *cay* and SLOPE. Strategy C uses *cay* and RREL. For full-sample *cay* lagged once, the manager runs regressions using quarterly time series at the end of each quarter as follows:

Strategy A:

$$R_t^e = \beta_0 + \beta_1 cay_{t-1} + \varepsilon_t \quad (2)$$

Strategy B:

$$R_t^e = \beta_0 + \beta_1 cay_{t-1} + \beta_2 SLOPE_{t-1} + \varepsilon_t \quad (3)$$

Strategy C:

$$R_t^e = \beta_0 + \beta_1 cay_{t-1} + \beta_2 RREL_{t-1} + \varepsilon_t \quad (4)$$

For recursive *cay* lagged twice, the manager uses $\widehat{cay}_{t-2}$ instead of cay_{t-1} in Equations (2), (3), and (4).

Finally, Strategy D is ex ante optimal from a statistical viewpoint. In each quarter, the manager runs predictive regressions with all possible combinations of predictive variables and chooses the forecast from the model that yields the lowest Schwartz criterion.

The manager uses least absolute deviation (LAD) regressions; this technique is more robust to outliers than ordinary least squares.[5] The excess return forecast is calculated using the latest value of the forecasting variables available. For example, in Strategy A it would be $\widehat{R^e_{t+1}} = \hat{\beta}_0 + \hat{\beta}_1 cay_t$ for full-sample *cay* lagged once, and $\widehat{R^e_{t+1}} = \hat{\beta}_0 + \hat{\beta}_1 \widehat{cay}_{t-1}$ for recursive *cay* lagged twice. Naturally, the estimated $\hat{\beta}$ coefficients are different in the two cases.

The manager also needs to know the volatility of the stock market's excess return to find the optimal portfolio. At the end of each quarter, the average volatility of the full sample of excess returns for daily data is used. Again, we assume that the information set consists of the history of returns until that time.[6]

Given the manager's risk aversion degree A, the weight to be allocated to the S&P 500 index at the beginning of each quarter is given by:

$$\omega_{t+1} = \frac{\widehat{R^e_{t+1}}}{A\hat{\sigma}^2_{t+1}} \tag{5}$$

In each of the two cases, we choose a risk aversion degree so that the average weight in the S&P 500 index across four strategies is equal to 1. We do this in order to facilitate comparison with a buy-and-hold strategy using only the S&P 500 index. It does not affect the most relevant performance evaluation techniques we use.[7]

The portfolio is formed at the end of each quarter and held until the next quarter. At the end of the next quarter, the manager uses new information to update the excess return forecasts, as well as to compute the new volatility estimate, and then rebalances each strategy's portfolio. We do not impose short-sale constraints.

PRELIMINARY RESULTS

Exhibit 1 plots the cumulative wealth under market timing strategies, the S&P 500 index, and the three-month T-bill. For example, one dollar invested in 1972 in the S&P 500 produces \$9.09 after 31 years, while Strategy C produces \$594.20 under full-sample *cay* lagged once and \$33.83 under recursive *cay* lagged twice.

Note that much of the improvement in performance is attributable to 22 quarters: 1973–1974 and 2000–2003:2. This could suggest that *cay* may be a more useful predictor of severe structural change than of high-frequency market movements.

All strategies not only end up above the S&P 500, but they are also always above it, with the exception of Strategy A under recursive *cay* lagged twice for a short period in the 1990s. Also note that full-sample cay lagged once strategies perform substantially better than recursive *cay* lagged twice strategies. (We discuss this difference later.)

Exhibit 2 presents the descriptive statistics for the share of wealth invested in the S&P 500 for all market timing strategies (choosing the risk aversion degree to make the average weight in the S&P 500 index equal to 1). Note that we allow for substantial leverage on some occasions. For example, Strategy C under full-sample *cay* lagged once might use up to four times the capital base. Naturally, most of our performance evaluation measures take this leverage into account.

Exhibit 3 tabulates the excess returns over three-month T-bills for each strategy. Note that the distribution of returns is more positively skewed than the market under all market timing strategies. This means that high positive excess returns are more likely than large negative excess returns compared to the S&P 500. Therefore, the mean-variance manager is not selling skewness to the market, a problem with the traditional approach to performance evaluation highlighted in Leland [1999]. Distributions also have fatter tails than the S&P 500 buy-and-hold strategy, however.[8]

One might argue that transaction costs could severely reduce the profitability of the market timing strategies, but that is not the case here. Assuming a 0.25% transaction cost per trade, a reasonable number for trading in futures contracts, we calculate that the average cost of following Strategy A under full-sample *cay* lagged once is 0.15% per quarter, very low compared to its mean excess return of 4.71% per quarter. Transaction costs for the other timing strategies are also very close to 0.15% per quarter. The effect is minimal because rebalancing occurs only quarterly.

STATISTICAL TESTS

First, we test the simple hypothesis that market timing strategies yield higher returns than the S&P 500. We compute the average difference between returns of the timing strategies and the S&P 500, and calculate the t-statistics of the difference (Exhibit 4).[9]

EXHIBIT 1
Cumulative Wealth over Time (log scale)

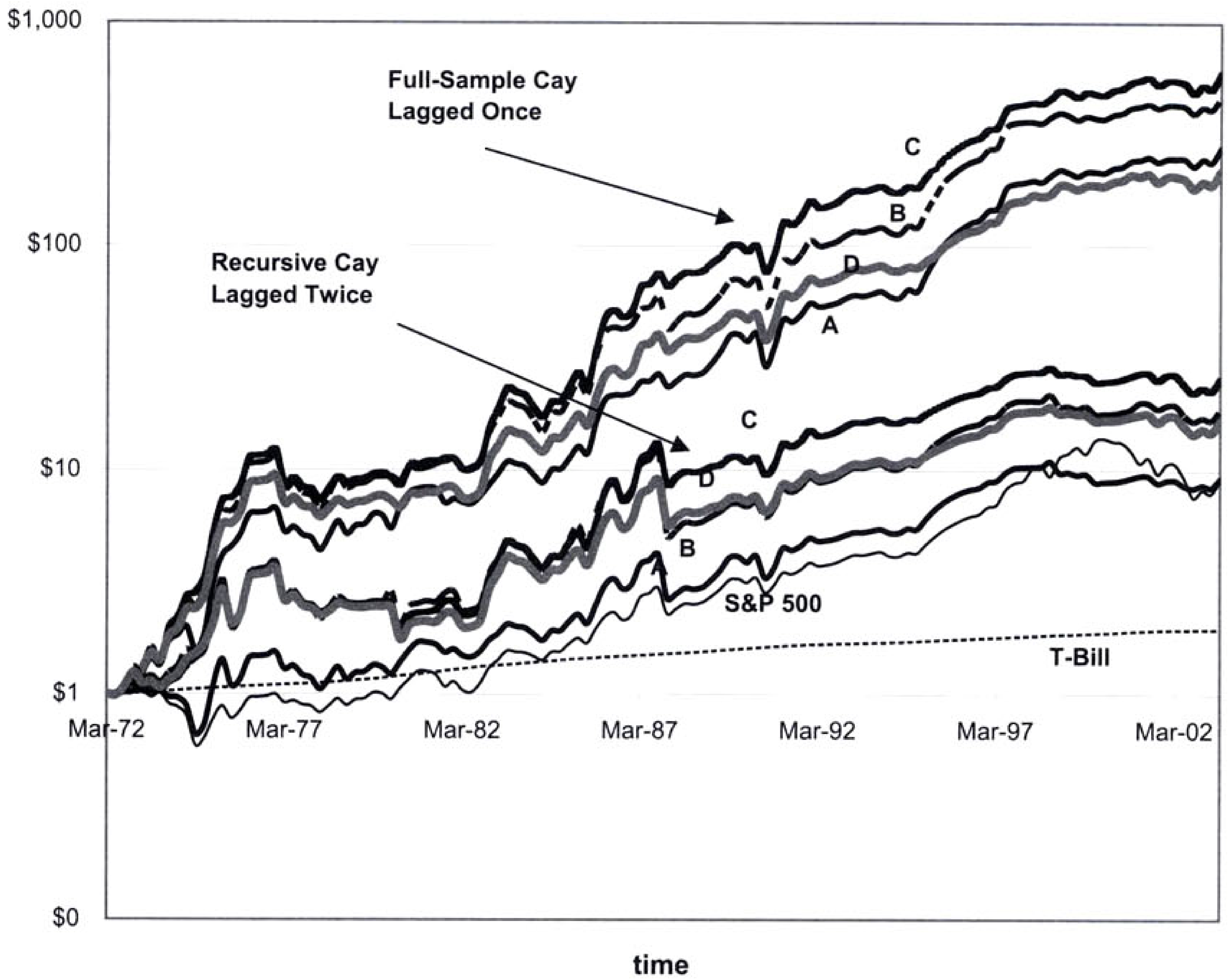

EXHIBIT 2
Descriptive Statistics of Portfolio Holdings

	Full-Sample *Cay* Lagged Once				Recursive *Cay* Lagged Twice			
	Mean	Std Dev	Min	Max	Mean	Std Dev	Min	Max
Strategy A	1.05	1.16	–2.99	3.24	0.98	0.63	–0.60	2.29
Strategy B	1.09	1.45	–4.38	3.86	1.09	1.35	–2.68	4.30
Strategy C	0.99	1.41	–3.96	3.37	0.99	1.12	–2.51	3.73
Strategy D	0.89	1.13	–2.57	3.37	0.94	1.06	–2.61	3.73

For all strategies the mean excess return over the S&P 500 is positive, but it is statistically significant for all strategies only under full-sample *cay* lagged once. (The compounded effect of the positive difference over 31 years can be seen in Exhibit 1.)

The t-test does not take into account the risk-return trade-off, so it is not suitable for assessing whether the manager is better off by following market timing strategies instead of a buy-and-hold strategy. That is, the t-test does not control for the fact that market timing strategies might require substantial leverage, and therefore are riskier than just buying and holding the S&P 500.

There are two ways to incorporate the risk-return trade-off into performance evaluation in the mean-variance world. The Jensen alpha test takes the view of a fully diversified investor who cares only about the systematic risk of an asset in the total portfolio. Sharpe ratio evaluation assumes investors place all their wealth in the

EXHIBIT 3
Descriptive Statistics of Quarterly Excess Returns

	Full-Sample *Cay* Lagged Once				Recursive *Cay* Lagged Twice			
	Mean %	St Dev %	Skew.	Kurt.	Mean %	St Dev %	Skew.	Kurt.
Strategy A	4.71	12.12	1.22	5.83	1.99	10.25	0.31	4.64
Strategy B	5.29	13.21	0.79	2.10	3.11	14.06	−0.11	5.28
Strategy C	5.46	13.11	1.15	2.86	3.07	12.63	0.65	3.37
Strategy D	4.42	11.19	1.08	3.08	2.61	11.48	0.34	3.66
S&P 500	1.60	8.47	−0.48	0.86	1.60	8.47	−0.48	0.86

EXHIBIT 4
T-Test of Difference of Quarterly Returns

	Full-Sample *Cay* Lagged Once	Recursive *Cay* Lagged Twice
	Return Net of S&P 500 (%) (t-stat)	Return Net of S&P 500 (%) (t-stat)
Strategy A	3.11 (3.77)	0.39 (0.67)
Strategy B	3.69 (3.92)	1.50 (1.48)
Strategy C	3.86 (3.94)	1.47 (1.63)
Strategy D	2.82 (3.55)	1.06 (1.20)

asset under consideration, and care about the total risk, not just the systematic part of it.

Exhibit 5 compares the Sharpe ratios of market timing strategies and the S&P 500 buy-and-hold strategy (Jobson and Korkie [1981]). All market timing strategies produce higher Sharpe ratios than the S&P 500 index, but the difference is statistically significant only for the full-sample *cay* lagged once. The Sharpe ratios of our strategies are independent of the degree of risk aversion. That is, we could choose a higher risk aversion A to make the strategies more conservative, requiring less leverage in either direction, without changing the Sharpe ratio results reported below.

Exhibit 6 reports estimates of the Jensen alpha of the timing strategies along with the t-statistic of the test that alpha is positive. All market timing strategies exhibit positive abnormal returns in the CAPM sense.[10] Moreover, alphas are statistically higher than zero for all strategies under full-sample *cay* lagged once, and close to significant for Strategy C under recursive *cay* lagged twice. In the first set of results, abnormal returns amount to roughly 13% per year, which is economically significant.

EXHIBIT 5
Quarterly Sharpe Ratios of Strategies

	Full-Sample *Cay* Lagged Once		Recursive *Cay* Lagged Twice	
	Sharpe Ratio	$SR_{str} - SR_{S\&P}$ (t-stat)	Sharpe Ratio	$SR_{str} - SR_{S\&P}$ (t-stat)
Strategy A	0.389	0.200 (2.56)	0.195	0.006 (0.09)
Strategy B	0.401	0.212 (2.55)	0.221	0.032 (0.39)
Strategy C	0.416	0.227 (2.59)	0.244	0.055 (0.67)
Strategy D	0.395	0.206 (2.54)	0.227	0.038 (0.47)
S&P 500	0.189	-	0.189	-

EXHIBIT 6
Quarterly Jensen Alpha

	Full-Sample *Cay* Not Lagged	Recursive *Cay* Lagged Twice
	Jensen Alpha (%) (t-stat)	Jensen Alpha (%) (t-stat)
Strategy A	3.22 (3.34)	0.49 (1.02)
Strategy B	3.78 (3.06)	1.52 (1.49)
Strategy C	4.07 (3.06)	1.62 (1.66)
Strategy D	3.10 (2.83)	1.30 (1.41)

Newey-West statistics with two lags.

It can be shown that although Jensen alphas are inversely proportional to the degree of risk aversion for our strategies, the t-statistics are invariant to the degree of risk aversion. For example, doubling the risk aversion degree A would cut Jensen alphas in half, without affecting the t-statistics. Again, we could increase the risk aversion coefficient in order to get less leverage in either direction, without affecting our conclusion about whether Jensen alphas are positive.

Henriksson and Merton [1981] devise a regression-based test to evaluate market timing ability. The test can be carried out with information about returns

EXHIBIT 7
Henriksson-Merton Market Timing Test

	Full-Sample *Cay* Lagged Once	Recursive *Cay* Lagged Twice
	C	C
	(t-stat)	(t-stat)
Strategy A	0.84	–0.03
	(1.48)	(–0.06)
Strategy B	0.96	0.09
	(1.92)	(0.12)
Strategy C	1.62	1.00
	(3.42)	(1.53)
Strategy D	1.29	0.64
	(3.18)	(1.05)

only, without knowing the data on portfolio holdings. The goal of the test is to see if the portfolio beta is higher when the market goes up than when it goes down. We run the regression and test whether coefficient c is positive:

$$R_t - r_t^F = a + b\left(R_t^M - r_t^F\right) + c\max\left[R_t^M - r_t^F, 0\right] \quad (6)$$

Exhibit 7 reports results for Henriksson-Merton test. All strategies produce positive coefficients with the exception of Strategy A under recursive *cay* lagged twice. All strategies under full-sample *cay* lagged once have at least a borderline statistically significant coefficient c. Only one strategy under recursive *cay* lagged twice has a borderline significant coefficient.

The Henriksson-Merton test requires data on portfolio returns only. As we have portfolio holdings in addition to returns, we are able to construct a more powerful test. We can use Spearman's rank correlation test to check whether weights in the S&P 500 index tend to be high when the subsequent excess return of the S&P 500 is high. This non-parametric approach is more robust to outliers than a linear correlation (or linear regression). The test uses the linear correlation ρ between the ranks of the series instead of the series themselves.[11]

Exhibit 8 presents t-statistics for the test that weights and subsequent excess returns are unrelated. Results indicate that, for all timing strategies, the correlation is at least borderline statistically significant, indicating that high weights tend to anticipate high excess returns on the S&P 500. Linear regressions also confirm this result.

It is also worth pointing out that the degree and the t-statistic of the Spearman correlation is invariant to changes in the coefficient of risk aversion. That is, leverage can be reduced without affecting this result.

Exhibit 1 reveals that returns in 1973–1974 are a major element in the high returns of the timing strategies. One might argue that structural changes in the economy could make early-period results less indicative of today's world. To address this issue, we examine the performance of the strategies during the latter half of the test period, 1988:1–2003:2. Note that this is a tougher test for the timing strategies, because it is always harder to achieve statistical significance in smaller samples, and also because we are removing a period in which we know, after the fact, that timing strategies did very well.

Exhibits 9 and 10 provide the results for this period for full-sample *cay* lagged once and recursive *cay* lagged twice. The subsample results are very similar to the full-sample results. Full-sample *cay* lagged once works well (results are positive and statistically significant for the most part), while recursive *cay* lagged twice produces weaker results (positive but not statistically significant). Note that Jensen alphas for recursive *cay* lagged twice are borderline significant in 1988–2003, as in the full sample.

EXHIBIT 8
Spearman Test: Holdings versus Subsequent Returns

	Full-Sample *Cay* Lagged Once	Recursive *Cay* Lagged Twice
	Spearman's ρ (t-stat)	Spearman's ρ (t-stat)
Strategy A	0.284 (3.27)	0.094 (1.05)
Strategy B	0.295 (3.41)	0.134 (1.50)
Strategy C	0.311 (3.62)	0.150 (1.69)
Strategy D	0.291 (3.66)	0.127 (1.41)

EXHIBIT 9
Summary of Quarterly Results Using Full–Sample *Cay* Lagged Once (1988-2003)

	Mean Excess Return %	Return Diff. % (t-stat)	Sharpe Ratio	SR Diff. (t-stat)	Jensen Alpha (t-stat)	Spearman ρ (t-stat)
Strategy A	4.25	2.12 (2.22)	0.456	0.181 (1.57)	2.63 (3.08)	0.223 (1.79)
Strategy B	4.14	2.02 (2.08)	0.450	0.175 (1.48)	2.59 (3.21)	0.174 (1.38)
Strategy C	3.75	1.63 (1.78)	0.424	0.149 (1.30)	2.20 (3.00)	0.102 (0.80)
Strategy D	3.07	0.95 (1.23)	0.389	0.114 (1.12)	1.54 (2.52)	0.097 (0.76)
S&P 500	2.13	-	0.275	-	-	-

As in 1972–2003, transaction costs are low. They reduce quarterly returns of the timing strategies by only 9 to 10 basis points per quarter assuming a 0.25% cost per trade. For Strategy A full-sample *cay* lagged once, for example, transaction costs would reduce the t-statistic of the t-test from 2.22 to 2.12, the Sharpe ratio from 0.456 to 0.445, and the t-statistic of the Jensen alpha from 3.08 to 2.96.

DISCUSSION OF RESULTS

We first discuss the performance difference between the two sets of results. In full-sample *cay* lagged once, we assume that in any quarter t our mean-variance manager knows Equation (1) for full-sample cay_t. This introduces some chance of a look-ahead bias, giving an unfair information advantage to our market timer.

EXHIBIT 10
Summary of Quarterly Results Using Recursive *Cay* Lagged Twice (1988–2003)

	Mean Excess Return %	Return Diff. % (t-stat)	Sharpe Ratio	SR Diff. (t-stat)	Jensen Alpha (t-stat)	Spearman ρ (t-stat)
Strategy A	2.78	0.66 (0.58)	0.302	0.027 (0.20)	1.59 (1.92)	0.125 (0.98)
Strategy B	2.95	0.83 (0.75)	0.334	0.059 (0.45)	1.82 (2.20)	0.027 (0.21)
Strategy C	2.37	0.25 (0.25)	0.275	0.000 (0.00)	1.04 (1.41)	–0.039 (–0.30)
Strategy D	2.31	0.19 (0.21)	0.286	0.011 (0.094)	0.98 (1.53)	–0.026 (–0.20)
S&P 500	2.13	-	0.275	-	-	-

In recursive *cay* lagged twice, the formula is reestimated each quarter, using only data available at that time. This removes the possibility of a look-ahead bias, but the use of reduced samples to estimate the cointegrating parameters necessarily yields noisier estimates of *cay*. These noisier estimates are likely to be further away from the true deviations of the long-run U.S. consumption-wealth ratio, and thus lead to inferior predictability and poorer market timing performance.

This problem is especially pronounced during early estimation recursions, where there are fewer data points. For example, when the formula is estimated using only data for 1951:4–1972:1, there are only 73 data points (after accounting for leads) but 37 parameters to be estimated. Therefore, both elimination of a look-ahead bias and higher sampling error could contribute to the poorer performance of recursive compared to full-sample estimates of *cay*, and there is no way to disentangle these effects.

In our results, the poorer performance also comes from using an additional lag for *cay* in the predictive regression. Results not reported, but available upon request, indicate that approximately 40% of the drop in performance comes from the additional lagging, while the remaining 60% comes from using recursive rather than full-sample estimation.

Why would we display the results with full-sample *cay* lagged once, when this strategy cannot be implemented in practice? We suggest two justifications.

First, suppose a money manager wants to start timing the market mechanically today using the strategies with *cay*. What Jensen alpha should the manager expect to get: an average of 3.5% per quarter as suggested by Exhibit 6 under full-sample *cay* lagged once, or 1.2% per quarter under recursive *cay* lagged twice? The answer is something in between.

In practice, the manager will use recursively estimated *cay* lagged twice, so the Jensen alpha should be no lower than 1.2% per quarter.[12] A manager starting with a much larger sample of data will not be forced to use rather unreliable estimates of *cay*, and should thus expect to have substantially better estimates of *cay* on average, i.e., closer to the true deviation from the long-run U.S. consumption-wealth ratio. Such a manager's Jensen alpha ought to be better than 1.2% per quarter, but probably less than 3.5% per quarter because of the potential look-ahead bias problem and the delay in availability of macro data. Therefore, the Jensen alpha (and other statistics) under full-sample *cay* lagged once serves to put an upper bound on performance.[13]

The second argument is conceptual. It may be possible to use other information to assess the current state of deviation from the long-run U.S. consumption-wealth ratio. We have confined ourselves to the cointegration relation developed by Lettau and Ludvigson [2001a]. Some investors may be more skilled in interpreting economic data available publicly, or may be able to get additional information about the U.S. economy from private sources.[14]

These investors, not constrained to mechanical use of public data, may be able to reap the full benefits of timing the market using true *cay*. The results with full-sample *cay* lagged once provide a better picture of the potential profits that a skilled market-timer might achieve.

Our full-sample period results with full-sample *cay* lagged once and recursive *cay* lagged twice suggest that augmenting *cay* with other predictive variables, such as the slope of the term structure and detrended short-term interest rates, may enhance market timing performance. It is worth pointing out that strategies relying only on SLOPE and RREL (without *cay*) fail to produce statistically higher Sharpe ratios than the S&P 500 or statistically positive Jensen alphas and Spearman correlation coefficients.

CONCLUSION

Results indicate that market timing strategies relying on true *cay*, the deviation from the long-run log consumption wealth ratio of the U.S. economy, deliver superior investment performance. Sharpe ratios increase from 0.19 to 0.40 per quarter, and Jensen alphas are economically and statistically significant. Results hold for both the full-sample period and the most recent subsample. Transaction costs are low and do not affect this conclusion.

To implement such strategies, however, requires estimating *cay* with available data and accounting for delays in macroeconomic data releases. When these real-world constraints are incorporated, results appear to be substantially weaker. The Sharpe ratios of the market timing strategies are still above the Sharpe ratio of the S&P 500, but the difference is not statistically significant. The Jensen alphas are positive but only borderline statistically significant. Results for the most recent subsample agree with those of the full sample period.[15]

Nevertheless, a trader who starts timing the market today may achieve better results since *cay* will be estimated from large samples. Moreover, macroeconomic data availability and quality are likely to improve over time.[16]

Also, traders might not be constrained to use macro data mechanically as our hypothetical managers are. For example, if proprietary research can assess the current state of the three variables that define *cay* more quickly and more accurately than the government, there are potentially large alphas to be gained. More research is needed to explore these possibilities.

The idea of successful market timing is consistent with Samuelson's [1994] view of financial markets: Although markets are nearly "micro"-efficient, in the sense of pricing securities relative to each other very well, they can be "macro"-inefficient, meaning that sometimes the overall level of asset prices does not rationally reflect all available information. Yet market timing is also consistent with a changing reward for bearing market risk over time as the result of an underlying equilibrium process.

ENDNOTES

The authors are grateful to the referee for many insightful comments and suggestions. They also benefited from useful discussions with Gregory Duffee, Guido Imbens, Jacob Sagi, and Mark Seasholes. They are thankful for the financial support of the Haas School of Business, the Dean Witter Foundation, and the CAPES Foundation.

[1]*Cay* data are quarterly and start in the last quarter of 1951. Lettau and Ludvigson [2001a] provide a detailed description of the estimation procedure for *cay*.

[2]This is similar to writing the consumption growth betas and the market risk premium as functions of *cay*.

[3]A more technical argument for the predictive power of the consumption-wealth ratio relies on cointegration of consumption and aggregate wealth as a result of an intertemporal budget constraint.

[4]Lettau and Ludvigson [2001a] use an equally weighted portfolio of all NYSE, AMEX, and Nasdaq stocks instead of the S&P 500.

[5]The results are similiar for OLS regressions.

[6]In general, finer sampling produces better volatility estimates. As we have daily data for the S&P 500 starting only in second-quarter 1963, we assume that the volatility over 1961:2–1963:1 is equal to the average volatility of 1963:2. As average volatility changes slowly over time, this is innocuous.

[7]This yields the risk aversion of 4.9 for full-sample *cay* lagged once and 4.2 for recursive *cay* lagged twice.

[8]Leland [1999] points out that since the aggregate investor also cares about skewness besides first and second moments, mean-variance managers could boost Sharpe ratios and Jensen's alphas by reducing skewness.

[9]The t-test makes sense here because the average weight on the S&P 500 is close to 1.

[10] Unconditional market betas are very close to 1.0 as a consequence of having the average weight on S&P 500 equal to 1.0.

[11] Admati and Ross [1985] argue that tests using portfolio holdings are potentially more powerful than those that use only returns. The Spearman test has also been applied in the context of timing ability by Chance and Hemler [2001].

[12] In fact, the manager can use a shorter lag, since macro data are available with a delay of less than one quarter.

[13] There is also an issue of revisions in macro data. According to Croushore and Stark [2000], macro data are revised several times after first publication, sometimes many years later, while our lagging in recursive *cay* lagged twice captures only the early revisions.

[14] Brokers, for example, have information about order flow for different assets.

[15] Relatively high Jensen alphas suggest that even market timing strategies based on recursive *cay* lagged twice might work as low-beta diversifiers in a fund of funds.

[16] Although the earlier availability of such data might just shorten the market adjustment time.

REFERENCES

Admati, Anat, and Stephen A. Ross. "Measuring Investment Performance in a Rational Expectations Equilibrium Model." *Journal of Business*, 58 (1985), pp. 1-26.

Campbell, John Y. "Stock Returns and the Term Structure." *Journal of Financial Economics*, 118 (1987), pp. 373-399.

——. "Understanding Risk and Return." *Journal of Political Economy*, 104 (1996), pp. 298-345.

Chance, Don M., and Michael L. Hemler. "The Performance of Professional Market Timers: Daily Evidence from Executed Strategies." *Journal of Financial Economics*, 62 (2001), pp. 377-411.

Croushore, Dean, and Tom Stark. "A Funny Thing Happened on the Way to the Data Bank." *Federal Reserve Bank of Philadelphia Business Review*, September/October 2000, pp. 15-27.

Henriksson, Roy D., and Robert C. Merton. "On Market Timing and Investment Performance. Statistic Procedures for Evaluating Forecasting Skills." *Journal of Business*, 54 (1981), pp. 513-533.

Jobson, J.D., and Bob Korkie. "Performance Hypothesis Testing with Sharpe and Treynor Measures." *Journal of Finance*, 36 (1981), pp. 889-908.

Leland, Hayne. "Beyond Mean-Variance: Risk and Performance Measures for Portfolios in a Nonsymmetric World." *Financial Analysts Journal*, January/February 1999, pp. 27-36.

Lettau, Martin, and Sydney Ludvigson. "Consumption, Aggregate Wealth, and Expected Stock Returns." *Journal of Finance*, 56 (2001a), pp. 815-849.

——. "Resurrecting the (C)CAPM: A Cross-Sectional Test when Risk Premia are Time-Varying." *Journal of Political Economy*, 109 (2001b), pp. 1238-1287.

Samuelson, Paul A. "The Long-term Case for Equities." *The Journal of Portfolio Management*, 21 (1994), pp. 15-24.