Universidad de San Andrés

Departamento de Economía

Trabajo de Graduación de Licenciatura en Economía

Return Predictability: the Campbell and Shiller Cyclically-Adjusted Price-to-Earnings Ratio

Background, Challenges and Empirical Evidence

Autor: Marcos Costantini
Legajo: 19051
Mentor: Daniel Aromí

Victoria, 2016

Table of Contents

Introduction

Originally presented in the academic context of stock return predictability, the Cyclically-Adjusted Price-to-Earnings ratio gained widespread used for broader stock market index valuation amongst practitioners with timely calls for disappointing stock returns following the late-1990's US equity bubble (driven primarily by technology company valuations). A whole host of literature subsequently emerged from Campbell and Shiller's original 1988 paper.

The aim of this paper is to serve as a survey on this ample literature concerning the Cyclically-Adjusted Price-to-Earnings ratio, in order to provide the reader with an overview of the academic discussions surrounding the model. We will focus our research on answering two main questions: first, whether it indeed shows empirical ability to predict subsequent stock returns, including in new out-of-sample data; and second, to analyze the range of explanations put forth to justify the existence of this anomaly and examine whether they are defensible.

The former is simpler to answer, while the latter encompasses several discussions, amongst them the role of risk and market efficiency, behavioral perspectives, questions surrounding non-stationarity of valuation ratios and demographic considerations. Our motivation for a survey-based approach lies in attempting to cover the full breadth of papers and discussions surrounding these two main research questions, which we feel have not been adequately covered in an individual paper before.

The rest of the paper is organized as follows. The first part will broadly attempt to answer the first research question. We will introduce the concept of market efficiency and provide a glimpse at the background on return anomalies in general, focusing on the value anomalies that act as a precursor to Campbell and Shiller's original work. We will then introduce the formal methodology of the Cyclically-Adjusted Price-to-Earnings (CAPE) ratio, its construction and the framework through which the original authors test the predictive power of long-term returns. Common statistical and methodological criticisms against the CAPE metric are examined, along with relevant counter-arguments. In addition, subsequent evidence supporting the validity of the CAPE ratio is presented, including research presented by academics and practitioners alike, including new out-of-sample data that was not part of the original model formulation.

The second part of the paper is comprised of theories that seek to explain this apparent return anomaly and covers the second major research question. As anticipated, this section covers several different arguments. We focus on authors that examine the matter of investment risk in the context of the Efficient Markets Hypothesis in order to explain differences in returns, as well as examine behavioral perspectives that sought to explain traditional return anomalies. Further, we will focus on the discussion of broad links between valuation ratios and demographics. Finally, we will cover papers that question the validity of a fundamental assumption – mean reversion in valuation ratios – underlying Campbell and Shiller's original work.

The third part deals with the most recent additions to the CAPE literature, which apply Campbell and Shiller's methodology to test stock return predictability with new data sets and methodological approaches. With the aim of covering innovations in the area, we present findings for the literature studying the CAPE anomaly internationally in emerging markets, as well as across market sectors and individual stocks in US equity markets.

The final part offers some conclusions on the extensive body of work surrounding the CAPE ratio and its potential usefulness in predicting long-term returns and offers suggestions for further research in the topic.

Part I: Empirical Return Anomalies and Campbell and Shiller's Cyclically-Adjusted Price-to-Earnings Ratios

A. Introduction to Market Efficiency and Background on Return Anomalies

Fama (1970) defines an *efficient* market as one in which prices fully incorporate available information. In doing so, the author introduces the concept of excess returns and models market efficiency in different ways[1]. Under the simplest *fair game* model, the expected future price of an asset at $t+1$ given a set of available information at period *t*, Ω_t, is today's price: $E\left[p_{t+1} \mid \Omega_t\right] = p_t$. Thus expected returns are $r_{t+1} = \mu + \varepsilon_{t+1}$, where $E[\varepsilon_{t+1} \mid \Omega_t] = 0$, implying non-predictability in asset price fluctuations.

Under these models, asset prices must be unpredictable because any reported anomaly that attempted to exploit information contained in Ω_t would quickly be arbitraged away by market participants, driving obtainable excess returns to zero. This conclusion shared by all models[2] and forms the basis of informational efficiency in markets.

Even prior to Fama (1970), considerable academic research focused on investigating the predictability of equity (stock) returns. Considerable evidence has been found to suggest the existence of a *value* anomaly: the tendency for stocks trading at the most attractive (cheap) valuations to significantly outperform their more unattractive (expensive) peers[3] (Basu, 1982; Fama and French, 1992; amongst others). This effect, initially shown on the US equities across several decades of data, was found to continue to be present on new out-

[1] Fama (1970) highlights the *fair game*, *martingale* and *random walk* models.
[2] The *martingale* model incorporates price and return time series and states that the expected future price of an asset, given an available information set, should – after discounting to the present – equal its current price. The *random walk* model for expected returns is a general extension of the previous models, and says that price changes across periods are independent and identically distributed, and that the entire return distribution is independent of the information set available.
[3] Refer to Appendix A for a full overview of this anomaly.

of-sample data and across international markets. Most importantly, these findings did not seem to be consistent with rational explanations of investment risk. For a full overview of these discussions, which are outside the scope of this paper, as well as an introduction to the relevant terminology of valuation metrics involved in these studies, please refer to Appendix A.

B. The Campbell and Shiller Cyclically-Adjusted Price-to-Earnings Ratio

Unlike prior research that broadly focused on *cross-sectional* stock return predictability, Campbell and Shiller (1988) set out to investigate the degree to which prices, accounting earnings and dividends can predict or explain subsequent real gross and excess returns over multiple timeframes for an *aggregate* stock market index. The initial motivation for the authors consisted in examining the degree to which stock returns were consistent with traditional present value models of dividends, which argue for a link between asset prices and discounted values of future dividends[4].

The data sets used are annual observations on prices, dividends and earnings for the Standard and Poor's Composite Stock Price Index, dating from 1871 to 1987, and deflated using the Producer Price Index annual series[5] in order to obtain real quantities and avoid the impact of inflation on returns altogether.

Campbell and Shiller define the real price of the stock index (measured at the start of the year *t*) as P_t and the real dividend paid during period *t* as D_t. The log gross return obtained by holding the index from the start of period *t* to the start

[4] The single-stage dividend growth model (also known as the Gordon Growth Model) accurately describes this relationship. Under this model, the value of a security is merely the present value of all future dividends in perpetuity, with excess earnings merely re-invested for future dividend growth (growth at rate *g*, discounted at *r*):

$$P = \frac{D_0(1+g)}{r-g}$$

[5] Known before 1978 as the "Wholesale Price Index" and compiled by the US Bureau of Labor Statistics (BLS).

of period t + 1 then becomes: $h_{1t} = \log\left(\frac{P_{t+1} + D_t}{P_t}\right) = \log(P_{t+1} + D_t) - \log(P_t)$. Thus, the real log total (price and dividend) return over i years (from t to $t-1$) obtained by a prospective investor is: $h_{it} = \sum_{j=0}^{i-1} h_{1,t+j}$

The authors also examine the predictability of excess returns over short-term interest rates[6], defining real log short-term interest rates in year t as r_t, and thus the excess return of stocks over i periods as: $h_{it} - r_{it}$.

The two regression models then become:

$$h_{it} = a + bX_t + \varepsilon \quad (1)$$
$$h_{it} - r_{it} = a + bX_t + \varepsilon \quad (2)$$

In their original paper, the authors regress both of these return metrics over multiple potential explanatory variables (X_t), including:

- The log dividend-price ratio, $\delta_t = d_{t-1} - p_t$
- The lagged dividend-growth rate, Δd_{t-1}
- The log earnings-price ratio (the reciprocal of the price-to-earnings ratio; using accounting earnings for the last year only), $\varepsilon_t = e_{t-1} - p_t$
- The ten-year moving arithmetic average of log real earnings minus current log real price, $\varepsilon_t^{10} = \frac{e_{t-1} + \ldots + e_{t-10}}{10} - p_t$
- The thirty-year moving arithmetic average of log real earnings minus current log real price, $\varepsilon_t^{30} = \frac{e_{t-1} + \ldots + e_{t-30}}{30} - p_t$

The author's motivation for including smoothed earnings series, as opposed to the traditional and more-often quoted one-year trailing earnings lies in the fact

[6] Calculated as the annual return of four to six month prime (investment grade) commercial paper, rolled over systematically in January and July.

that earnings for a particular year could potentially be excessively more volatile than the underlying worth of the companies themselves. They use the rationale set forth by Graham and Dodd (1934), who argued for security selection on the basis of cyclically-adjusted, or *smoothed* earnings over a full business cycle, highlighting the volatile nature of one-year earnings that can be heavily influenced by external factors such as macroeconomic shocks and social influences, and are furthermore compounded by the degree of total operating and financial leverage employed by the businesses. Indeed, taken to the extreme, one-year earnings could theoretically be negative whereas equity values, as call options on residual claims of the business, must by definition have a non-negative value.

The choice in the number of years for which cyclically-adjusted (smoothed) earnings are computed in Campbell and Shiller's paper, ten and thirty years, partly exceeds Graham and Dodd's recommended choice of seven to ten years of earnings data; the author's rationale corresponds to the fact that decadal variability of earnings, in their view, likely exceeds the variability in underlying fundamental value for the shares themselves, such that the thirty years could also prove an appropriate timeframe for the study.

The authors find that the variables with predictive power in their regression study for the sample of S&P 500 from 1871 to 1987 are the log dividend-price ratio, δ_t, and the three earnings-to-price ratios: log earnings-price, ε_t, and the smoothed 10-year and 30-year log earnings-price ratios, ε_t^{10} and ε_t^{30}.

The authors find statistically significant forecasting power for these variables, with the cyclically-adjusted (smoothed) log price-earnings variables proving better predictors than the log dividend-price ratio, the lagged dividend-growth rate and the log earnings-price across all time horizons (refer to Figure 1). Though the variability explained by all of these variables is low at short-term horizons – between 1.9% and 6.7% for all four variables at a 1-year horizon, for example – the forecasting power of these variables increases in every instance as the time horizon progressively expands from one to ten years: the smoothed 10-

year and 30-year log earnings-price explain 40.1% and 56.6% of the subsequent 10-year real absolute return variability, respectively. Similar results are obtained when considering excess real returns over short-term debt interest rates, with these same variables explaining 34.1% and 48.0% of real excess returns at the 10-year timeframe.

Campbell and Shiller's conclusion holds true whether one analyzes discounted or undiscounted returns: a significant degree of subsequent returns are predictable using cyclically-adjusted earnings series, such that high initial levels of smoothed price-to-earnings ratios are correlated with poor subsequent returns over long horizons, and vice versa. The relationship using 30-year smoothed real earnings is depicted in Figure 2. It is worthy to point out that excess returns over short-term interest rates are found to be just as strongly correlated with initial log earnings-to-price ratios as absolute returns, showing that it is not the forecastability of interest rates by market participants that drives this relationship. For these reasons, the smoothed 10 and 30-year log earnings-price ratio and its non-log counterparts, the 10-year and 30-year cyclically-adjusted price-to-earnings ratio (hereafter referred to as CAPE10 and CAPE30, respectively), become the predominant smoothed P/E metrics in the literature.

Predicting Stock Returns, 1871–1987[a]

	Exact Returns (expression 1, text)			Discounted Returns (expression 4, text)		
	1-year	3-year	10-year	1-year	3-year	10-year
			A. Real Returns			
Explanatory variables						
δ_t	0.039	0.110	0.266	0.048	0.135	0.327
	(0.033)	(0.015)	(0.001)	(0.017)	(0.006)	(0.000)
Δd_{t-1}	0.000	0.004	0.003	0.000	0.004	0.003
	(0.964)	(0.522)	(0.485)	(0.977)	(0.568)	(0.537)
ϵ_t	0.019	0.090	0.296	0.023	0.104	0.303
	(0.143)	(0.027)	(0.000)	(0.100)	(0.017)	(0.000)
ϵ_t^{10}	0.040	0.111	0.401	0.047	0.130	0.423
	(0.036)	(0.031)	(0.000)	(0.022)	(0.019)	(0.000)
ϵ_t^{30}	0.067	0.195	0.566	0.079	0.225	0.615
	(0.013)	(0.008)	(0.000)	(0.007)	(0.004)	(0.000)
δ_t, Δd_{t-1}, ϵ_t^{30}	0.076	0.204	0.637	0.088	0.235	0.667
	(0.073)	(0.046)	(0.000)	(0.041)	(0.022)	(0.000)
			B. Excess Returns			
δ_t	0.016	0.080	0.184	0.022	0.101	0.246
	(0.180)	(0.037)	(0.033)	(0.114)	(0.019)	(0.010)
$\Delta d_{t-1} - r_{t-1}$	0.026	0.027	0.000	0.026	0.026	0.001
	(0.082)	(0.127)	(0.811)	(0.082)	(0.134)	(0.758)
ϵ_t	0.011	0.054	0.195	0.015	0.066	0.206
	(0.261)	(0.083)	(0.009)	(0.194)	(0.053)	(0.005)
ϵ_t^{10}	0.052	0.145	0.341	0.060	0.168	0.399
	(0.017)	(0.010)	(0.003)	(0.010)	(0.005)	(0.001)
ϵ_t^{10}	0.051	0.187	0.480	0.074	0.218	0.548
	(0.017)	(0.007)	(0.002)	(0.009)	(0.003)	(0.000)
δ_t, $\Delta d_{t-1} - r_{t-1}$, ϵ_t^{30}	0.086	0.195	0.493	0.096	0.229	0.553
	(0.046)	(0.045)	(0.011)	(0.028)	(0.022)	(0.004)

FIGURE 1: Results of 1, 3 and 10-year return regressions on log dividend-price, lagged dividend-growth ratio, log earnings-price and average 10 and 30-year log earnings-price ratios, from Campbell and Shiller (1988). "Exact Returns" corresponds to expression 1 in this text, while "Discounted Returns" corresponds to expression 2. The numbers reported are the R^2 of the regression; the numbers in parenthesis are significance levels of a Wald test hypothesis that all regression coefficients are equal to zero.

A revised study is conducted by Shiller (1996), who revisits the topic in order to address two issues: on the one hand, the author updates prior findings using new data made available since the original study (encompassing the period from 1988 to 1996); and secondly, the author addresses potential statistical issues surrounding the original paper[7]. Indeed, by incorporating the additional earnings, price and return data from 1988 to 1996, the R^2 in the original regression for 10-year real absolute returns on the ratio of the smoothed 30-year log earnings-price ratio (CAPE30) rises further to 0.624 (statistically significant at all confidence intervals). The evidence for 10-year real discounted returns is just as strong as in the original formulation, as the R^2 for the regression remains largely unchanged. Further work by Campbell and Shiller (2001) aimed to once again evaluate prior findings with new data following the severe US equity market correction in 2000. The authors find that the explanatory power of the CAPE10 on subsequent 10-year stock returns rose to 40% by including all data available up to the year 2000.

Subsequent work by the authors attempted to directly address their prior findings in light of the Efficient Markets Hypothesis (EMH): Campbell and Shiller (1998), under the assumption that valuation ratios are long-term mean-reverting, argue that when valuation ratios themselves (in this case dividend-to-price and price-to-earnings) become extreme relative to history, then either the ratios must predict subsequent dividend or earnings growth (the nominator in the equation), or should well be predictors of subsequent price performance (the denominator). The authors make the point that, under the traditional EMH model, stock prices (and consequently returns) should not be predictable on the basis of publicly available financial information (from which valuation ratios are derived). Therefore, in their view, for the EMH to hold, as the dividend-to-price and price-to-earnings ratios run at extreme values relative to history, then the ratios must be predictors of subsequent growth in dividends and earnings.

[7] The main statistical issue addressed by Shiller (1996) is the potential bias in estimated coefficients given stochastic regressors in the original regression. For further details, refer to section 1.E.

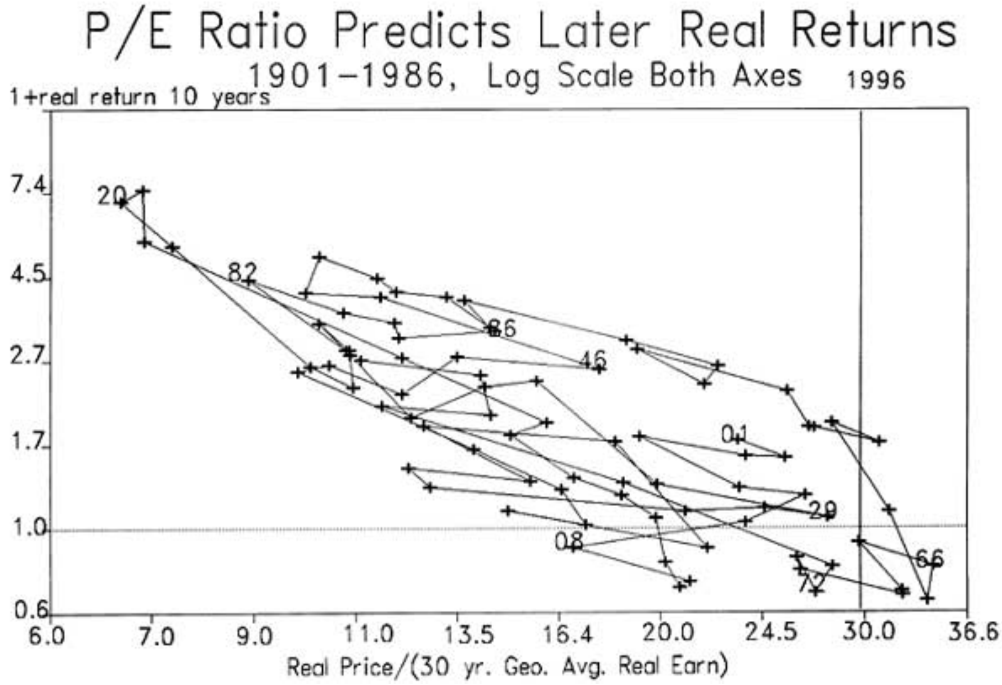


FIGURE 2: Plotting subsequent 10-year relative returns vs. starting 30-year log earnings-price ratios (R^2 of 62.4%), from Shiller (1996).

To this effect, the authors regress price growth and earnings growth (over 1 and 10-year horizons) versus the initial value of the smoothed 10-year price-to-earnings (non-log) ratio. The find poor predictive ability for the 10-year price-to-earnings ratio to forecast future growth in smoothed earnings (R^2 of less than 4% over 1-year and 10-year timeframes), and equally poor performance for the dividend-to-price ratio to forecast future growth in dividends. On the other hand, in support of their prior papers, they find that the 10-year smoothed price-to-earnings ratio is a solid forecaster of subsequent 10-year growth in stock prices, with R^2 of 37%. The authors conclude that for both the dividend-to-price and price-to-earnings ratios, it is the change in the price component of these ratios that is largely responsible for bringing these measures back to long-term averages, not subsequent growth in dividends and earnings. Therefore, Campbell and Shiller's findings stand in apparent contradiction to the concept of informational efficiency in asset prices, as widely-available historical information – such as price-to-earnings ratios – should not allow an investor to obtain consistent positive excess returns.

C. Persistence of Predictor Variables

A major point of argument on the statistical front for empirical research that, like Campbell and Shiller, study the predictability of long-term returns on lagged variables has been the debate regarding the inference issues that arise in regression studies that have regressors with high degrees of persistence (near unit root or unit root).

In such a scenario, even granting the key assumption that the regression's explanatory variable (smoothed price-to-earnings ratio) is stationary and thus mean reverting, utilizing traditional *t*-tests and critical values in these regressions may trigger Type 1 errors under which the null hypothesis of non-predictability may be falsely rejected at traditional *t*-test critical values by virtue of persistence in the regressors alone. This argument, put forth by Stambaugh (1986) and Mankiw and Shapiro (1986) show the inherent bias towards predictability in most of the traditional literature studying the signal value of earnings-to-price, book-to-market, dividend-to-price and similar metrics. Nelson and Kim (1993) and Stambaugh (1999) make the case that biased coefficients are expected when the predictor variables themselves are strongly correlated with stock returns. Given the fact that these predictor variables (valuation ratios) are by construction a function of non-stationary asset prices (Roll, 2002), the problematic is evident.

We are exclusively concerned with the potential effect of this bias in the original Campbell and Shiller regression[8]. To this effect, Campbell and Yogo (2006) find sufficient persistence in the valuation ratios to determine that the traditional *t*-statistic test findings of predictability in long-term returns to be biased. The authors build on similar work by Lewellen (2004) and Torous, Valkanov and Yan (2004) and construct a new *t*-test that is more robust to persistence in the

[8] A debate on the validity of the traditional literature (using non-smoothed time series) studying predictability of stock returns given these findings is beyond the scope of this paper. For a more complete background on the topic and research into the value anomaly after implementing methodological changes to account for these issues, refer to Lewellen (2004) and Campbell and Yogo (2006).

predictor variable, and can be utilized in standard regression methods[9]. To this effect, they implement a new regression with data for the 1926 to 2002 period, finding that the 10-year smoothed price-to-earnings ratio is still capable of predicting returns at various intervals (monthly to annually) after accounting for these issues, supporting the significance of Campbell and Shiller's original findings.

D. Out-of-sample Performance

Striving to re-examine the performance of models that attempt to predict equity returns on the basis of valuation metrics, Goyal and Welch (2006) argue that many of the previously-reported 'anomalies' are plagued by model issues that lead them to be unstable or spurious, leading to non-significance even with in-sample data used in the original regressions. They further criticize the out-of-sample performance of these models, making the case that most of them have poor predictive ability and would have likely been unable to help investors in allocation decisions on a real-time basis. This problem, the author argue, is not isolated to any decade of potentially outlier empirical return evidence, but is indicative of structural problems that plague such models. In the particular case of the CAPE10 model performance on the US market, Goyal and Welch find statistically significant in-sample performance for the 1927 to 2004 period, but they conclude the model has no statistical value in out-of-sample testing[10]. In contrast, they conclude that historical returns themselves are overall better predictors of future returns.

Campbell and Thompson (2008) argue against Goyal and Welch's (2006) findings for poor out-of-sample performance of valuation predictor variables: they run new regression studies on the full S&P 500 earnings and returns series

[9] The authors use a Bonferroni procedure based on a modified Dickey-Fuller unit root test and Q-test in order to generate a more robust confidence interval for the coefficient for the dependent variable.

[10] The authors test all 1-year, 3-year, 5-year and 10-year smoothed price-to-earnings ratios configurations with monthly, yearly and 5-year frequency, finding limited out-of-sample performance even in the best cases, with particularly poor performance in the 3 prior decades (1974-2004).

from 1872 to 2005 with two main modifications: the first one involves simple sign restrictions on the regression coefficients, such that coefficients with a value that is opposite of the expected sign is truncated to zero; the second modification involves non-negative equity premiums (that is, the intercept in the regression must be equal to or larger than 0). While Campbell and Thompson found Campbell and Shiller's 10-year smoothed price-to-earnings ratio economically meaningful in predicting annual returns in and out-of-sample without imposing restrictions on the regression, the modifications proposed further enhance the performance of the model in and out of sample, and improve upon Goyal and Welch's previous criticism regarding the inconsistency in predictive power from the CAPE10 across the different sub-periods in the data (noting the weak performance in the 1974-2004 subset, for example). Campbell and Thompson further show how even relatively small improvements in the R^2 of the regression can have dramatic benefits for investors concerned with the risk-adjusted performance of timing models, further concluding that the benefits of investing on the basis of CAPE10 vis-à-vis a static buy and hold strategy can be economically significant[11].

E. Biases in Stochastic Regressors

A potential statistical issue concerning the original CAPE model findings is addressed by Shiller (1996), who highlights that a certain degree of bias in the estimated coefficient is expected given stochastic regressors in the original regression in Campbell and Shiller (1988). The author argues that, assuming earnings are sufficiently smoothed over sufficiently long period (eg: 10 or 30 years), the cyclically-adjusted price-to-earnings ratio may display mean-reversion tendencies towards the sample mean, even in the scenario where stock prices were completely decoupled from earnings[12].

---

[11] The authors run backtests under which equity allocation can vary between 0% and 150% (leverage implied) according to CAPE10 levels: they find a minimum and maximum of 44 and 191 basis point annual outperformance across the worst (1956-1980) and best (1927-1956) subset periods, respectively, once these conditions are imposed on the regression coefficients.

[12] Quoting Shiller (1996): "In simple terms, even if stock prices have no relationship at all to simple earnings, so long as earnings are smoothed enough to generate the price–earnings ratio, there will tend to be a negative correlation small samples between the price earnings ratio and

To this effect, the author runs a Monte Carlo simulation under which 10-year price returns follow a random walk, leaving 10-year smoothed earnings as a constant throughout the sample period. 10-year returns on the random walk time series are regressed against the initial smoothed earnings value. Shiller finds that the realized R^2 of the original regression, with a value of 0.624, only occurs in less than 2% of simulations, indicating a high degree of statistical significance.

The author subsequently repeats the experiment using 30-year lagged index prices as a proxy for lagged 30-years earnings data (rationalized by highlighting the strong historical fit between the two series) yielded similar conclusions: only about 0.26% of simulations yielded R^2 values in the range of those observed in the results for the original regression.[13]

F. Discussion of Methodology Behind the Construction of the Cyclically-Adjusted Price-to-Earnings Ratio

There have been a number of authors – notably Siegel (2013) and Wilcox (2011) – who have criticized the contemporary utility of the original Campbell and Shiller Cyclically-Adjusted Price-to-Earnings ratio in predicting future returns. Broadly speaking, these authors argue that as a result several key changes in the methodology of the individual components that make up the CAPE ratio, the original formulation has become all but meaningless in evaluating the true underlying state of cyclically-adjusted earnings for an equity index and, consequently, in predicting future equity returns.

The first major argument presented by Siegel (2013) emphasizes how recent changes in accounting methodology have made the present-day earnings figures

the thirty-year average of earnings. The negative correlation arises primarily because the sample mean is estimated over the whole sample, and prices will naturally appear to be mean reverting to their sample mean, even if no true mean exists."

[13] Campbell and Shiller (1989) run Monte Carlo simulations under which returns – and not prices – follow a random walk pattern. The results prove consistent with their other findings.

that go into the denominator of the price-to-earnings ratio incompatible with historical values. The author notes the wide divergence between reported operating income and those of GAAP earnings[14] for S&P 500 firms during the past two United States downturns in 2002-2003 and 2008-2009, an unusual occurrence that did not occur in previous recessions. Siegel claims that these events occurred as a result of accounting changes: he cites FASB 144/142 on impairments of intangibles ("write-downs" on values of intangibles depress earnings and cannot be revalued upwards later should conditions improve, thus biasing earnings downwards over time) and treatment of "available for sale" securities by financial institutions (to always be reported at fair market value). These changes, the author concludes, have made current stated earnings incompatible with historical earnings, thus distorting the data that goes in to the CAPE metrics and therefore diminishing its usefulness as a predictive metric for subsequent returns[15]. Montier (2014) argues that historical earnings figures have had multiple accounting changes over time, including changes that have historically adjusted earnings upwards during down cycles. He uses the case of FASB 157 (suspension of mark-to-market on financial company assets) in 2009 as a regulatory change that skewed stated earnings in the opposite direction (upwards). The implicit argument Montier makes is that accounting changes over time have not always uniformly biased earnings in one direction and that one should expect them to cancel out over time. Asness (2012) makes the broader case that the popular one-year trailing price-to-earnings and book-to-market ratios that market practitioners commonly cite (particularly relative to history) are equally susceptible to changes in accounting standards, such that the potential problem is not only limited to smoothed metrics such as Campbell and Shiller's CAPE10.

[14] Generally Accepted Accounting Principles (GAAP) is the standard accounting framework under which US Corporations report their financial statements, as set out by the Financial Accounting Standards Board (FASB). Stated GAAP Earnings are the official net income figures for US corporations and are therefore used in the calculation of earnings for indices such as the S&P 500. It is S&P 500 GAAP earnings that go into the denominator of the CAPE metric.

[15] Siegel proposes changing S&P Reported Earnings in the Campbell and Shiller CAPE for NIPA profits (National Income and Product Accounts, measured across 9,000 companies), suggesting improved forecasting power as a result of the elimination of said earnings distortions driven by write-downs. Hester (2013) argues against such change, noting NIPA metrics track earnings from current production and exclude capital gains/losses, making them an imperfect proxy for S&P earnings as a whole.

The second major argument put forth by Siegel refers to what he calls *aggregation bias* in index earnings construction, that becomes especially problematic when the reported earnings of the firms with the highest weightings in the index have exceptionally poor performance. This skews earnings for the whole index downwards in times of stress[16]. Siegel argues that the aggregation methodology of S&P earnings misstates earnings power of the other index components[17]. In times of severe stress, Siegel argues, index earnings are systematically under-reported, and thus the CAPE may lose its predictive ability and anticipate a more depressed return going forward.

Finally, Siegel also maintains that there has been a secular shift in the financing decision of firms which led them to decrease dividend payout ratios and increase retention ratios, thereby increasing long term growth rates[18] and systematically leading to seemingly-inflated CAPE values given lower earnings in early years (for instance, years 1-3 for the CAPE10), an underlying change that Campbell and Shiller's CAPE construction cannot compensate for. To examine the impact of the rate of earnings per share growth on the final CAPE metric, consider the example of two separate 10-year periods with differing growth metrics: in the first case, real earnings per share rise linearly from \$50 to \$100 over the 10 years, while in the second case, real earnings steadily rise from \$80 to \$100 over the period. For the former case, the ending CAPE would be constructed on the basis of the average of real earnings throughout the 10-years, in this case, \$75 in real earnings per share. In the latter case, the average would be \$90 in per share real earnings. Thus, the ending CAPE10 value would be significantly different in the two examples assuming the same index price (i.e.: the first and second cases would have a CAPE10 reading of 13.33 and 11.11, respectively, with the index priced at \$1000 at the end of the 10-year period). It becomes evident that the

[16] For instance, negative S&P500 net earnings in the fourth quarter of 2008 as a result of massive write-downs by major firms such as AIG, Citigroup and Bank of America.

[17] Taken to the extreme, if one assumes these stocks are worthless, the earnings impact on the index should only be proportional to the company's weight in the index: they cannot have negative earnings values that subtract from earnings in other index components.

[18] Assuming a company's theoretical long-term growth rate (*G*) is its Return on Equity (ROE) times its Retention Ratio (or capital that is not distributed to shareholders, but is rather re-invested in the business and generate returns equivalent to the company's internal return on equity):

G = ROE x (1 – Retention Ratio)

rate of growth in real earnings is fundamentally important in determining the value of the CAPE given its impact in the smoothed earnings series. To be sure, the purpose of smoothed earnings is to discount more aggressively very rapid increases or decreases in real earnings, which can be heavily influenced by the stage of the business cycle and may thus be temporary in nature. However, Siegel makes the case that firms themselves can influence the rate of growth in earnings over time as a consequence of their capital allocation decisions (by decreasing dividend payout ratios, for instance), and that such decisions will tend to bias the CAPE higher relative to history and reduce its efficacy as a market timing tool.

The choice of 10 years of historical earnings data that goes into the construction of the original CAPE10 metric is criticized by Wilcox (2011), noting economic cycles in the United States have historically been closer to 6 years. Thus, the author argues, the choice of timeframe is incorrect in smoothing out the economic cycle since it encompasses an average of roughly 1.5 historical economic cycles. The author also notes the differences in Consumer Price Index (CPI) calculation by the Bureau of Labor Statics in the United States across time[19], which had the effect of lowering the reported inflation rate over time. The argument is that this creates additional problems when comparing CAPE ratios (which are by definition inflation-adjusted) over multi-decade periods. He finally mentions the same accounting changes that Siegel pointed out and their impact on the comparability of CAPE over time. Asness (2012) and Faber (2013) note the general inconsequential effect of using 10 years of data in the "traditional" CAPE10 calculation, pointing to prior research by Butler and Philbrick (2011), who find similar forecasting power with a CAPE constructed of varying time periods (from 2 to 30 years), leading to the view that the choice of timeframe itself is not important. Rather, the authors maintain, the act of smoothing out regular fluctuations in year-to-year earnings when computing valuation ratios is the important point of Campbell and Shiller's contribution, not the choice of any particular timeframe.

---

[19] For example, by introducing hedonic models to account for changes in quality, which would shift CPI downwards, all else being equal.

As a result of some of these arguments, Bunn and Shiller (2014) extend the original CAPE ratio and identify methodology improvements in order to address two key criticisms presented against it: the authors modify the ratio to eliminate the bias that varying inflation calculations may present and eliminate the impact that changes in corporate payout policy over time have on the CAPE ratio.

To be clear, it bears recognizing that both Siegel (2013) and Wilcox (2011) believe in the validity of the original anomaly as presented by Campbell and Shiller (1988); however, for a variety of reasons highlighted above, they have become skeptical of the present utility of the original CAPE formulation in predicting future equity returns and hence favor modifications to the original methodology for it to regain acceptable use as a market timing tool.

G. Recent Evidence on Predictive Ability

There has been considerable recent literature that attempted to replicate and extend Campbell and Shiller's analysis as per the original model's formulation. Most research has focused on three key aspects: examining the explanatory power of the cyclically-adjusted price-to-earnings ratio on the original Campbell and Shiller (1988) regression using more recent data[20] (the time elapsed since the 1987 data adds at least two non-overlapping 10 year data points); analyzing the profile of future return distributions across initial cyclically-adjusted price-to-earnings figures; and extending the original analysis to encompass return analysis across a higher number of timeframes [21].

One of the first financial market practitioners to test the validity of the original Campbell and Shiller study with more recent data is Faber (2010). The author incorporates data available up to the end of 2010 and finds an R^2 of 33% for a regression of 10-year real returns vs. log of starting 10-year cyclically-adjusted

[20] Campbell and Shiller performed several revisions to their original paper, all of which are highlighted in Section 1.B. This section consists on findings by other authors working with a more expanded dataset and applying new analysis to their findings.
[21] While Campbell and Shiller (1988) originally studied future returns at 1, 3 and 10-years timeframes, the authors covered here expand the timeframes considerably.

price-to-earnings ratios, in-line with Campbell and Shiller's original results. Faber also finds significant differences in subsequent 10-year real returns depending on the starting CAPE10 level: by splitting the data into 10 initial CAPE10 deciles (from the lowest CAPE readings in 1920 to the highest in 2000 for the US equity market) and computing subsequent returns, the author finds differences in returns of close to 10% (annualized) between top and bottom deciles. Asness (2012) finds similar results, and additionally exposes significant differences in the best and worst case scenarios for realized returns across starting CAPE10 deciles (refer to Figure 3). Blanchett, Finke and Pfau (2013) agree with previous literature as to the statistically significant nature of CAPE on future 10-year real returns (R^2 of 23.8% and t-statistic of -7.65, significant at all levels), while the predictive power for shorter timeframes declines to 6% for 1-year forward returns, albeit still statistically significant according to their results, with a *t*-statistic of -2.88 (significant at a 99% confidence level).

Results For S&P 500 From Different Starting Shiller P/Es 1926-2012

Starting P/E Low	Starting P/E High	Avg. Real 10 Yr Return	Worst Real 10 Yr Return	Best Real 10 Yr Return	Standard Deviation
5.2	9.6	10.3%	4.8%	17.5%	2.5%
9.6	10.8	10.4%	3.8%	17.0%	3.5%
10.8	11.9	10.4%	2.8%	15.1%	3.3%
11.9	13.8	9.1%	1.2%	14.3%	3.8%
13.8	15.7	8.0%	-0.9%	15.1%	4.6%
15.7	17.3	5.6%	-2.3%	15.1%	5.0%
17.3	18.9	5.3%	-3.9%	13.8%	5.1%
18.9	21.1	3.9%	-3.2%	9.9%	3.9%
21.1	**25.1**	**0.9%**	**-4.4%**	**8.3%**	**3.8%**
25.1	46.1	0.5%	-6.1%	6.3%	3.6%

FIGURE 3: Return distribution from starting CAPE10 deciles. Source: Asness (2012)

Similar work is done by Goddard (2011), who analyzes returns at shorter timeframes and finds significant differences in outcome distributions for average 3-year subsequent real returns for US stocks on the basis of initial CAPE10 levels: the author identifies a 10.4% annualized delta in returns between the cheapest and most expensive initial quartile (as measured by CAPE10). Goddard also finds significant differences in frequency of negative returns on the basis of initial CAPE10: no negative occurrences over subsequent *aggregate* 3-year period in the cheapest quartile vs. 32% of occurrences in the most expensive quartile. Braun and Kaussen (2014), following the previous literature and extending the Campbell and Shiller dataset with data up to 2013, find a strong negative correlation between starting CAPE10 (sorted by quintiles) and subsequent 10-year forward returns, with the delta between the most expensive and least expensive starting CAPE quintiles on 10-year forward return of 11.7% per year, in line with previous findings.

One of the first studies that considerably extended the timeframes used in return predictability in the CAPE framework was done by Butler and Philbrick (2011): covering different starting CAPE configurations (with cyclically-adjusted earnings that vary from 2 to 30 years) and returns across varying time periods (5 years, 10 years, 15 years, 20 years and 30 years), the authors find that the explanatory power of the smoothed price to earnings ratios generally improves as the time horizon is extended, in support of Campbell and Shiller's original conclusions, with predictive power peaking in the 20-year forward return range. Unlike prior literature, however, the authors conclude cyclically-adjusted price-to-earnings ratios for the US markets have no statistically significant forecasting ability at time horizons of 5 years and shorter. Brodeski, Beall and Larson (2012) use data available through 2010 and similarly look at predictive power at longer timeframes: the authors analyze the relationship between initial CAPE10 figures and subsequent 1 year, 5 year, 10 year and 20 year real total returns, finding correlation and explanatory power rise as time horizon is extended, with an R^2 for the regression of 30.8% and 57% and correlations of 0.56 and 0.75 for 10-Year and 20-Year subsequent real returns, respectively. They also find an 8% and 6.3% annual real return delta over 10 and 20-years, respectively, between the highest initial CAPE10 quartile and the bottom CAPE10 quartile throughout the sample period.

To this effect, we find general agreement from both academics and market practitioners alike as to the validity of Campbell and Shiller's original work, particularly on its effectiveness at predicting returns at longer timeframes (above 5 years). The authors note that the out-of-sample data made available since the model's inception in 1988 did not serve to disprove its validity[22].

H. Relationship Between Macroeconomic Variables, Cyclically-Adjusted Price-to-Earnings Ratio and Return Predictability

There have been a number of authors that attempt to examine the relationship between the market's prevailing valuation level and the state of macroeconomic variables. Due to the fact that the market's (conventional 1-year trailing) price-to-earnings ratio can be notoriously volatile[23], these authors set out to investigate the degree to which the market's "normalized" valuation ratio (in this case represented by CAPE10) is influenced by macroeconomic considerations.

One of the first authors to study this topic was Arnott (2011), by examining the relationship between the market's prevailing CAPE10 ratio, real interest rates[24] and inflation[25]. The author finds a strong relationship between CAPE10 and these two variables: there is evidence for a Gaussian-like distribution for the CAPE10, which peaks in value in the 2-3% inflation and 3-4% real interest rate range, decreasing steadily as inflation and real rates rise or decrease substantially from that point (see Figures 4 and 5). Arnott further finds a 68% correlation between historical CAPE10 values and those predicted by this two-variable model of real interest rates and inflation[26], suggesting a high degree variability in investors' willingness to pay more or less for (smoothed) equity earnings can be explained by real rates and inflation figures.

[22] This subsection covers studies by academics and market practitioners alike. While the latter group's papers – amongst which one can highlight Faber (2010), Goddard (2011) and Butler and Philbrick (2011) – may be viewed somewhat more skeptically given the lower potential degree of research rigorousness (given the lack of peer reviews and so forth), it bears reminding that all authors are broadly in agreement as to the conclusions derived.
[23] Primarily due to earnings volatility as a result of cyclical forces. For a full discussion on the theoretical justification for cyclically-adjusted earnings series, refer to Section 1.A.
[24] Calculated by subtracting trailing 3-year (annualized) CPI from the nominal yield on the 10-year Treasury bond.
[25] Trailing 3-year (annualized) CPI inflation.
[26] With a significant t-statistic of above 7 after adjusting for overlapping samples inherent in the CAPE10 methodology.

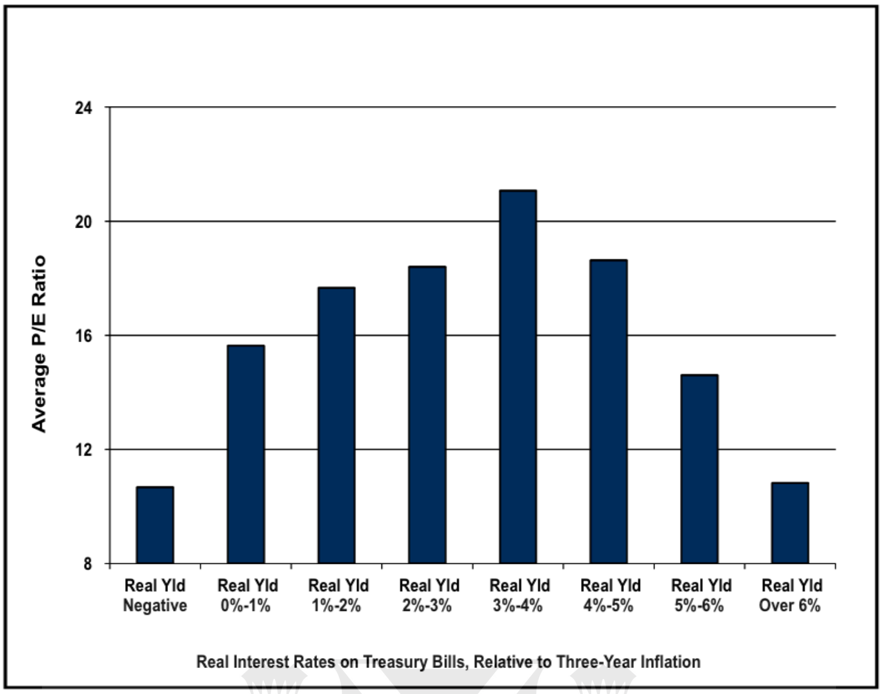


FIGURE 4: Relationship between Shiller CAPE10 Ratio and Real Interest Rates for US Market, 1871-2010. Source: Arnott (Research Affiliates, 2011)

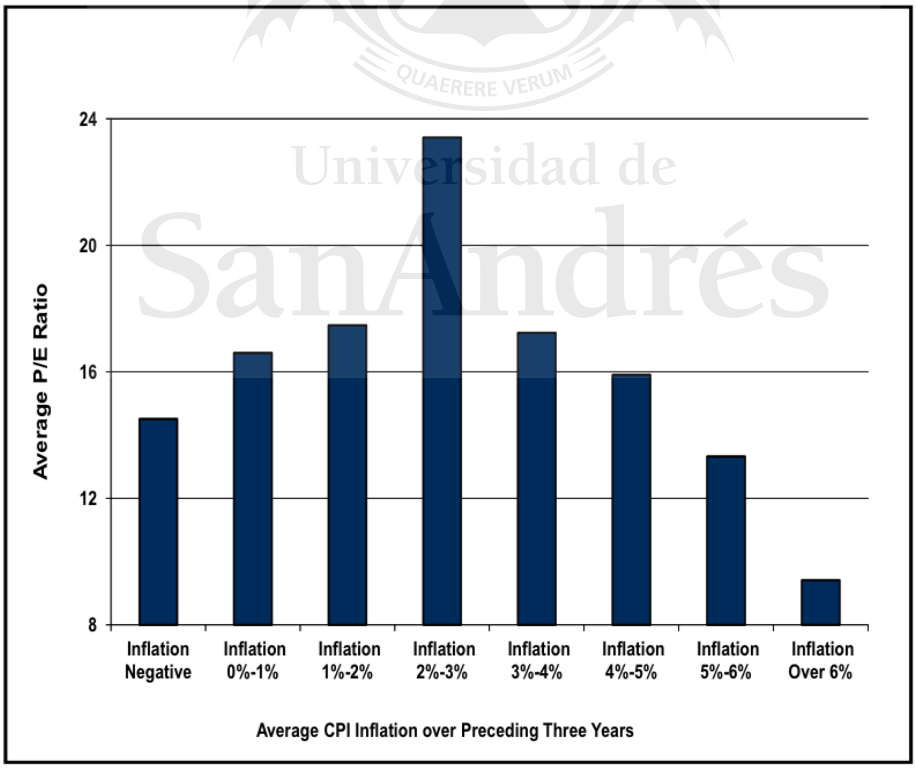


FIGURE 5: Relationship between Shiller CAPE10 Ratio and Trailing 3-year CPI Inflation for US Market, 1871-2010. Source: Arnott (Research Affiliates, 2011)

Studying the significant impact of changes in valuation ratios on investor returns over 5-10 year timeframes[27], Parikh (2012) finds an additional relationship between prevailing CAPE10 ratios, realized nominal GDP growth and volatility in nominal GDP growth. The author's conclusion, in line with Arnott (2011), is that there is an optimal combination of macro variables that results in higher investor willingness to pay higher multiples for a given quantity of earnings: nominal GDP growth in the 2% to 6% range results in highest valuations, while growth beyond and below these levels are not only associated with higher GDP growth volatility, but also have idiosyncratic effects, since faster growth has been associated with rising interest rates and slower growth with depressed earnings environments.

Both author's findings suggest that investor's willingness to pay different prices for equal amounts of equity earnings (expressed via varying cyclically-adjusted price-to-earnings ratios) may be the result of a non-linear relationship with the state of key macro variables. To be clear, both Parikh (2012) and Arnott (2011) are studying the relationship between macroeconomic variables and the market's historical valuation (as measured by CAPE10) in these different scenarios, not the relationship between these macroeconomic variables and realized returns.

On this latter point, Davis, Aliaga-Díaz and Thomas (2012) study many previously-proposed explanatory variables for future real returns at a stock market level (S&P 500 in the US from 1926 to 2011), including macroeconomic variables, and find most of the often-cited metrics (trend of real GDP growth, 10-year Treasury yields, expected future real GDP growth, and profit margins for the corporate sector, for example) to be poor forecasters of future returns. Davis, Aliaga-Díaz and Thomas had four main conclusions: first of all, out of the 15 potential variables analyzed in the study, the CAPE10 was the variable with the

[27] Parikh finds that over extremely long time horizons (in the order of 100 years), returns obtained by investors in equities are primarily obtained via contributions from growth (driven by nominal GDP growth and the share of corporate profits as a percentage of the economy) and income (dividend payouts and share buybacks), while the return contribution from changes in valuation is "essentially zero".

most explanatory power of future 10-year real returns throughout the sample (R^2 of 0.43); secondly, the choice of smoothing period is largely irrelevant, with cyclical adjustments between 2 and 15 years performing similarly (peak forecasting power in the 8-10 year range); thirdly, the extreme valuation outliers in the data (for example, extremely high CAPE10 levels in the 1990s that correctly predicted lackluster future returns) are found not to be responsible for the model's high degree of forecasting power[28]; and lastly, that forecast ability is meaningful only above 5 year timeframes, in line with previous findings.

[28] Davis, Aliaga-Díaz and Thomas found that only by dropping a full 25% of the sample (12.5% of peak and 12.5% of trough values) did the R^2 drop to 0.20 for 10-year future real returns

Part II: Questioning the Validity of the Campbell and Shiller Cyclically-Adjusted Price-to-Earnings Anomaly

A. Rational Explanations of Risk and Risk-Adjusted Return Evidence in Model Implementation

Fama (1970) succinctly summarized the prevailing attitude from academics towards the Efficient Markets Hypothesis at that time: "In short, the evidence in support of the efficient markets model is extensive, and (somewhat uniquely in economics) contradictory evidence is sparse." The author's basic principle was that markets were quite efficient in pricing in all new information regarding individual stocks and the overall market, such that it would be unfeasible for an investor to systematically exploit strategies that incorporated past stock prices or return data (in the *weak form* EMH) or public fundamental data (in the *semi-strong* form), such as earnings or dividends, to obtain returns that exceed a random portfolio of stocks (once the risk of the strategy and real costs such as trading commissions are taken into account).

Subsequently, Fama and French (1992) found evidence that the size and value factors[29] offer significant explanatory evidence for the cross-sectional dispersion of security returns in US equity markets from 1963 through 1990. The evidence suggested that Sharpe's Capital Asset Pricing Model (1964), which used the security's exposure to systematic risk (beta), did not fully explain returns. This went in direct contradiction to the model as it was presumed that bearing exposure to systematic risk was the only source of risk that compensated investors (whereas idiosyncratic security risk did not). The authors further suggest that the excess returns earned by high book-to-market firms might be merely rewards that investors receive for bearing other risks, given the evidence for low book-to-market firms to display "persistently weak" economic

[29] The size factor attempts to measure a given company's market cap size versus its peers, while the value factor used a company's relative Book-to-Market ratio (inverse of Price-to-Book ratio) versus peers.

performance (the opposite is suggested of high book-to-market firms)[30], and the fact that high book-to-market metrics predict lower future earnings, and serve as compensation for higher distress risk (Fama and French, 1995). Similar rationale is employed when justifying the superior performance of small firms, echoing Chan, Chen and Hsieh (1985) that argue for the firm size variable acting as a proxy for fundamental economic risk factors. Chen and Zhang (1998) draw similar conclusions, while Aretz, Bartram and Pope (2007) suggest that book-to-market, size and momentum factors are all proxies for and provide exposure to macroeconomic risk factors[31]. Fama and French (1992) further highlight Ball's (1978) conclusion that variables in the traditional anomaly literature (size, leverage, book-to-market) are merely the inverse of a scaled stock price and therefore proxies for variables that track expected returns (earnings-to-price and dividends-to-price).

Surprisingly few arguments from the efficient market viewpoint have been explicitly provided to rationalize the evidence presented in Campbell and Shiller (1988). To some extent, this is undoubtedly because it becomes difficult to address the apparent CAPE anomaly, which involves returns series for the entire stock market, using arguments levied against cross-sectional return anomalies for individual stocks. After all, the scope of the problem changes: when dealing with the aggregate stock market, justifications of idiosyncratic company risk being captured by the proxy variables of valuation ratios do not apply. Nonetheless, the arguments put forth by these authors can be made against Campbell and Shiller's findings in a similar fashion: that is, that the high expected returns at times in which the stock market exhibits unusually low cyclically-adjusted price-to-earnings ratios can either be compensation for other types of risks (*systematic* rather than *idiosyncratic* risk), or they can merely respond to changes in discount rates.

The latter point is directly tackled by Campbell and Shiller (1988) in their original study. On the one hand, they run the same regressions on excess stock

[30] Fama and French (1992)
[31] Including inflation, the interest rate curve, growth expectations and exchange rates.

returns (relative to short-term corporate interest rates), finding equal conclusions about the significance of cyclically-adjusted earnings. Secondly, they propose a dynamic version of Gordon's dividend growth model[32] and, by generating expectations running a vector autoregression system based on the assumption of constant expected real returns on stocks, the authors find that stock prices and returns – while highly correlated with theoretical values – are significantly more volatile over the short-term than the model would suggest. These results suggest stocks have been prone to significantly higher return volatility than justified on a fundamental basis (considering changes in growth and discount rates).

Regarding the separate issue of high expected returns during periods of low starting CAPE values as serving as compensation for bearing risks – identifiable or otherwise – a number of authors have attempted to address the topic through empirical studies. We will first cover the question of identifiable risks[33], focusing on the traditional concept of financial risk: volatility of returns.

Significant research has focused on studying the risk profile of returns derived from trading models that allocate to equities on the basis of cyclically-adjusted price-to-earnings ratios. The argument goes as follows: if the conventional rational explanation justifying higher returns due to increased levels of risk (return dispersion) proves to be correct, one would expect to see no significant empirical evidence of excess risk-adjusted returns when implementing systematic strategies that use cyclically-adjusted price-to-earnings ratios to determine equity allocation exposure. In this subset of the literature, risk was traditionally defined as *ex-post* realized volatility (or dispersion) of returns, with

[32] According to the Gordon Growth Model, $P = \frac{D}{r-g}$, or $\frac{D}{P} = r - g$, where D/P is dividend-to-price, r is the (constant) discount rate and g is the long-term dividend growth rate. The authors modify the model such that it allows the dividend-to-price ratio to vary across time as a result of changes in the discount rate and/or growth rate. Dividend-to-price should therefore be tied to the present value of expected one-period interest rates and dividend-growth rates (Campbell and Shiller, 1988).

[33] Behavioral considerations and "unidentifiable" risks are covered in Section 2.B.

the resulting Sharpe ratio[34] providing for direct risk-adjusted return comparability. Given some of the inherent weaknesses in equity return volatility (and its ensuring effects on the Sharpe ratio)[35], some authors further identify downside volatility[36] and maximum drawdowns[37] of said strategies as ancillary risk metrics.

As we will examine, the methodology itself is varied: some authors test a system under which an investor establishes 100% equity allocations in periods in which the market is below its rolling long-term mean, and replaces for bonds or cash in periods in which the market is deemed more expensive than its long-run average; while other authors scale equity exposure according to the degree of over or undervaluation (relative to historic values at that point). Regardless, the common theme throughout this literature involves using the Campbell and Shiller cyclically-adjusted price-to-earnings ratio (CAPE) to construct active portfolios with varying exposure relative to the passive buy-and-hold approach, much like the traditional value anomaly literature tested active strategies by buying *value* and selling *glamour* stocks. The key difference however, is that the former authors are only concerned with varying equity exposure to a market index (such as the S&P500) across time relative to an always-invested buy-and-hold approach on the basis of valuation ratios applied to the entire stock market, while the traditional value anomaly literature[38] was concerned with security selection and exclusion on the basis of company-specific valuation ratios.

---

[34] The Sharpe ratio provides a handy measure for comparing risk-adjusted returns for different strategies, and measures the return obtained *in excess* of the risk-free rate per unit of volatility (higher is better): $S = \frac{R_p - R_f}{\sigma_p}$, where R_p, R_f and σ_p denote the strategy (or portfolio's) CAGR, the risk-free rate and the strategy's realized volatility, respectively.
[35] Volatility metrics assume a normal distribution for equity returns, which has historically not been the case, as equity returns for major markets have shown to exhibit negative skew and leptokurtosis (Sheikh and Qiao, 2009). Further, the Sharpe ratio becomes economically insignificant when portfolio returns are below the risk-free rate.
[36] Downside volatility, also known as downside deviation, refers to volatility of returns when said returns fall below the minimum acceptable return (MAR) threshold (typically 0%). In this way, the volatility of periods with positive returns is excluded (as it is suggested investors would be less concerned with volatility in periods in which they are making positive returns).
[37] Maximum drawdowns refer to peak-to-trough performance in periods of negative returns.
[38] Refer to Appendix A for further background.

Long-term trading models using Shiller CAPE are investigated by Faber (2011), who covers available data from 1900 through 2010. Using a simple model that is 100% invested in equities when CAPE levels are below (rolling) average CAPE (otherwise 100% cash), re-examined annually, Faber finds similar gross returns to buy-and-hold (9.93% CAGR[39] for the model vs. 9.41% for buy-and-hold), with substantially lower volatility (14.4% model vs. 20% buy and hold), thereby improving Sharpe ratios from 0.27 to 0.42 and reducing maximum drawdowns from the index's 65.2% to 25.3%. Faber also examines similar models built on arbitrary cut-off points (such as CAPE below 15, or 20), and while the findings are in favor of the trading model, the rolling CAPE mean model is highlighted to avoid accusations of data mining. No considerations for transaction costs or tax impacts are considered.

Pfau (2011a) and Pfau (2011b), addressing previous findings of Fisher and Statman (2006)[40], finds CAPE10 values provide a solid framework for trading signals built on rolling historical median values (above which the portfolio shifts towards bonds and below which stocks are favored), earning similar compounded returns to buy-and-hold but with substantially lower volatility (both total and downside volatility) and lower drawdowns, in effect providing investors with higher risk-adjusted returns than pure buy-and-hold (Fisher and Statman's ideal choice of benchmark). Pfau also compares the market timing system towards a more balanced 50%/50% stock and bond mix (present in Fisher and Statman's methodology) finding significantly higher returns at comparable risk levels. Pfau repeats the experiment with varying trading model allocation rules (from more balanced towards stocks to more balanced towards bonds), to avoid data mining bias accusations, finding similar results throughout:

[39] Compound Annual Growth Rate (CAGR) is merely the geometric average of annual returns, and is an annualized figure.

[40] Fisher and Statman (2006) examine the usefulness of dividend-to-price, price-to-earnings, cyclically-adjusted price-to-earnings (over multiple time periods) and investor sentiment figures in market timing, without consideration for the risk profile of said strategies. The authors conclude that most intuitive trading rules – those that in the authors' views do not depend on extensive data mining – were deemed inferior to a simple buy-and-hold approach for US stocks (as represented by the S&P500) in the period 1964 to 2002. Braun and Kaussen (2014) issue similar conclusions by examining stated (not risk-adjusted) returns. The lack of consideration of investment risk in Fisher and Statman's (2006) study forms the basis for Pfau (2011a) and Pfau (2011b), who highlights the importance of looking at risk-adjusted returns.

risk-adjusted returns are positive when incorporating trading rules built on trailing cyclically-adjusted price-to-earnings ratios. Solow, Kitces and Locatelli (2011) find simple allocation changes at extreme values, when smoothed price-to-earnings ratios are at the top or bottom deciles of broad backward-looking historical values can lead to excess risk-adjusted returns (higher Sharpe ratios) for investors relative to a simple buy-and-hold approach.

Butler and Philbrick (2013) extend the CAPE methodology by subtracting latest inflation rate from the inverse of the CAPE (that is, the smoothed earnings yield) to arrive at the market's real cyclically-adjusted earnings yield. They propose different valuation-based asset allocation models that incorporate this metric. The first one invests 100% of assets in S&P 500 when the valuation is below the (rolling) long-term average, and otherwise holds all cash, examining results from 1934 through 2012. They find that while the this approach produced lower CAGR[41] (7.5% vs. 10.5% for pure buy-and-hold), it does so with almost half the volatility (6.1% vs. 13%), resulting in higher Sharpe ratio for the trading model vs. buy and hold (0.61 vs. 0.51), as well as significantly lower peak drawdowns of 13.2%, vs. 49% for buy-and-hold. The second model invests in stocks only when the S&P 500's real earnings yield is above the (rolling) 80th percentile, and otherwise holds cash. Results are even better, with CAGR of 11.4%, volatility of 9.6%, Sharpe of 0.8 and peak drawdowns of 26.3%, though the choice of 80th percentile will undoubtedly seem arbitrary and lead to suggestions of data mining. To this effect, Butler and Philbrick examine a model in which the percentage allocated to equities matches the percentile of the market's current earnings yield relative to history (on a rolling basis), with the rest assigned as cash[42]. This eliminates the need for the investor in question to select an admittedly arbitrary cut-off point between stocks and cash, as was the case in the second model. This model produces similar results as the first one, with CAGR of 7.7%, volatility of 5.1%, Sharpe of 0.77 and maximum drawdown of 17%. Similar results are obtained when Treasuries are considered in lieu of cash for the un-invested portion. One potential criticism of the methodology is the

[41] In agreement with prior literature.

[42] For instance, if the current cyclically-adjusted real earnings yield is at 75% percentile on 'cheapness' on a backwards looking basis, the system allocates 75% to equities and 25% to cash.

high cost of mimicking the strategy prior to the introduction of index funds, particularly in the third model that re-balances monthly, as well as the potential relative tax-inefficiency of the strategy.

Regardless of which metric is taken as a definitive proxy for investment risk – be it return volatility, drawdowns or downside volatility – these authors have separately come to broad agreement that trading strategies devised on CAPE have historically provided consistently higher risk-adjusted returns relative to a passive approach[43], which suggests markets have been consistently over-compensating participants willing and able to bear these risks, and undermining the notion that identifiable systematic risks are responsible for Campbell and Shiller's original anomalous findings.

This naturally leads one to consider other types of risks of an unidentifiable nature, the so-called "model risks". The argument could be made that historical periods in which the market's CAPE ratios were extremely low coincided with instances of radical uncertainty and unquantifiable catastrophic risks that – while evidently did not occur in the case of US equity markets – does not automatically negate their existence as a potential outcome[44]. By definition, these risks are unquantifiable and are therefore beyond the limits of quantitative research, but they are nonetheless valid arguments to consider. We would nonetheless note that such "model risk" instances seem to be mainly associated with negative tail risks rather than positive tail risks, and therefore the exceptionally poor returns associated with extremely high initial CAPE ratios would not likely be fully addressed by this framework.

[43] This subsection covers studies by market practitioners. There is understandably very little in the way of academic studies on backtested performance of CAPE models. While said results may be viewed somewhat more skeptically given the lower potential degree of research rigorousness (given the lack of peer reviews and so forth), it bears reminding that all authors are broadly in agreement as to the conclusions derived.

[44] While not addressing this point specifically, Faber (2015) provides two enlightening instances of countries whose equity markets lost 100% of their value amidst deep political and economic transitions: Russia in 1917 and China in 1949.

B. Behavioral Explanations for the CAPE Anomaly

Thus far the focus has lied on rational explanations of risk – identifiable or otherwise – to justify the apparent CAPE anomaly. There is another area of discussion, which is succinctly introduced by Pfau (2011b): after finding convincing evidence for positive risk-adjusted returns from trading models using CAPE ratios, the author deems the primary obstacle against achieving said results to be of a *behavioral* perspective[45]. We would note that there have been few direct arguments put forth by behavioral economists regarding the rationale behind Campbell and Shiller's findings, and would instead turn to arguments put forth in the context of the traditional return anomalies[46].

The original arguments for evidence of "overreaction" from investors came from DeBondt and Thaler (1985), who found that portfolios consisting of recent underperformers tended to consistently gain higher risk-adjusted returns than those formed by recent outperformers, suggesting investors "attach disproportionate importance to short-run economic developments". Another prominent argument from the behavioral family is the issue of overconfidence in extrapolating future growth. Lakonishok, Shleifer and Vishny (1994) find evidence for outperformance of strategies that buy out-of-favor (*value*) stocks relative to glamour (*growth*) stocks. Most importantly, they find that future realized growth rates of the glamour stock group were materially lower than they had been in the past, indicating a potential extrapolation of past trends well into the future. While these authors offered behavioral explanations for these apparent anomalies present in cross-sectional studies, a similar point is made on the general equity market by Campbell and Shiller (1998), who highlights the tendency for valuation ratios (smoothed or otherwise) to be elevated in periods in which the economy has exhibited strong growth, and suggests investors may become overly complacent when the recent trends in growth and earnings have been wholly positive (and vice versa).

[45] As opposed to a *cost* perspective: given the absence of low-cost index funds for the majority of the sample, replicating said strategies in prior decades might have been costly in terms of transactional and accompanying tax costs.
[46] These are outlined in Appendix A.

Shiller (2000) makes these points as he examines the tendency for markets to swing from boom to busts, often times driven by speculative bubbles whose feedback mechanism involves many of the same psychological patterns that behavioral economists previously identified: treating most recent prices as fair levels for a stock or stock market, following Kahneman and Tversky's (1974) observation of an anchoring bias; overconfidence and expectational feedback loops, based on work by Barberis, Shleifer and Vishny (1998); herd behavior and information cascades, following Bikhchandani, Hirshleifer, and Welch (1992); illusion of control and overconfidence in selective outcomes, based on Langer (1975); and herding based on social pressures, following Asch (1952). Shiller (2000) makes the argument that all the previously-mentioned cognitive biases work together to help amplify the feedback loop through which panics and manias drive valuations for the equity market. The evidence for mispricings above and below fundamental value is developed by Shiller (1981), who finds evidence for aggregate stock market price volatility to be significantly higher than that justified by volatility in underlying fundamentals.

From the empirical results derived from CAPE trading models (as outlined in Section 2.A.), the high excess expected returns during periods of excessively low cyclically adjusted price-to-earnings ratios could not have been rewards to investors for bearing investment risk (in the traditional sense), as by definition risk-adjusted metrics already account for these considerations. Therefore, excluding the issue of unidentified risks – which by definition are highly uncertain and escape quantitative analysis – one may conclude that Campbell and Shiller's anomaly may be explained by behavioral reasons under which risk aversion systematically decreases in boom periods and increases in busts (i.e.: markets overshoot their fundamental value in both directions), leading to trading opportunities for investors willing to take a contrarian approach[47]. In this sense, the high excess returns might well serve as compensation to investors

[47] Given the fact that results from these papers suggested that conventional financial risks (as defined by volatility of returns) involved in these types of strategies were, if anything, subtler than a simple buy-and-hold approach. This applies both to conventional risk metrics (volatility), as well as to other complementary figures, such as downside volatility and maximum drawdowns. Refer to Section 2.A. for further background on these topics.

for bearing psychological risks by going against all natural behavioral biases: perhaps the clearest counter-example is that of an asset manager that, afraid of losing clients due to underperformance in strongly positive markets, increases equity exposure despite overwhelming evidence that stocks are expensive on all fundamental metrics.

To be clear, there were undoubtedly heightened risks during periods in which valuations were depressed relative to history (1930's depression and early 1980's inflation, for instance) but the issue is whether these risks were overstated, for behavioral reasons, such that expected returns were *too* high. It seems that, given the fact that the CAPE model testing finds consistent evidence of excess returns after controlling for investment risk, the conclusion may well be that either unidentified risks or behavioral biases must be largely responsible for these results at an aggregate stock market level: while the former is by definition impossible to test, the latter has been present across cross-sectional stock results and serves as a valid explanation for these findings.

C. Non-Stationarity of Valuation Ratios and the Fed Model

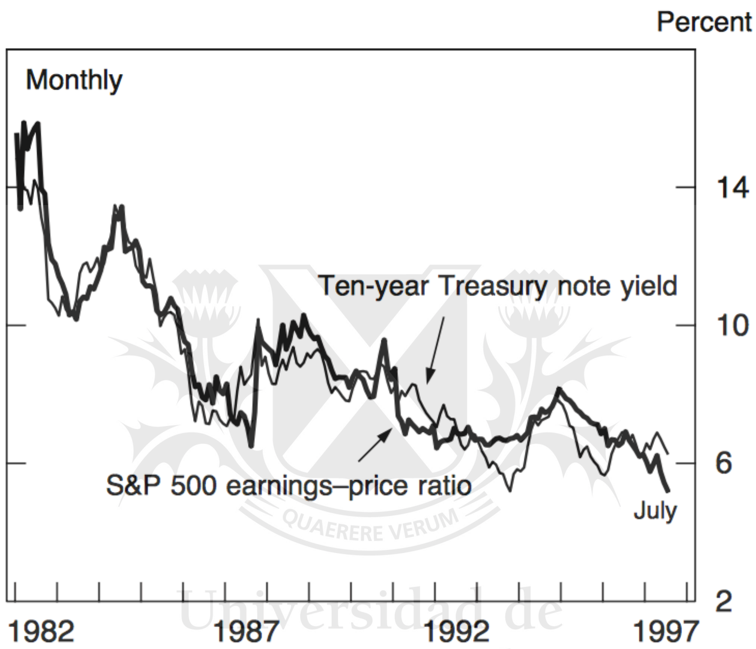
The original Campbell and Shiller (1988) model and the host of supporting literature implicitly assume that the market's price-to-earnings ratio (cyclically-adjusted or otherwise) is a stationary – and thus mean-reverting – time series, by virtue of being a linear combination of two cointegrated variables (earnings and prices). Campbell and Shiller (1998) make this point explicitly, as they state: "If we accept the premise for the moment that valuation ratios will continue to fluctuate within their historical ranges in the future, and neither move outside nor get stuck at one extreme of their historical ranges, then when a valuation ratio is at an extreme level either the numerator or the denominator of the ratio must move in a direction that restores the ratio to a more normal level". By subsequently disproving the forecastability of earnings and dividend growth on the basis of initial price-to-earnings and dividend-to-price ratios, respectively, the authors find a significant statistical link between initial valuation ratios and future returns, within the framework containing the working assumption of stationary valuation ratios.

Weigand and Irons (2006) investigate this underlying assumption behind Campbell and Shiller's work by comparing it to the so-called "Fed Model"[48]. This data, as shown in Figure 6, depicted the close relationship between the yield on 10-Year US government Treasury bonds and the US stock market's real earnings yield (the reciprocal of the price-to-earnings ratio) from 1982 to 1997. The interpretation of the data, and the main proposition behind the "Fed Model", is that the market's earnings yield will tend to gravitate towards nominal interest rates (represented by the yield on the 10-year Treasury bond), as this represents the benchmark to which returns are compared against: "In its simplest form, [the Fed Model] asserts stocks are cheap when [the market's earnings yield] exceeds [10-year yields], expensive when [10-year yields] exceed [the market's earnings yield] and fairly valued when [both] are equal" (Asness, 2003). The model essentially suggests investment flows will flow to the asset class (stocks or bonds) with the most attractive yield at any given time. The Fed Model itself, while criticized for its simplistic and incorrect theoretical basis, has been deemed useful in helping determine how investors use it in pricing the market's price-to-earnings ratio (Yardeni, 2003; Malkiel, 2004), as emphasized by the high degree of correlation between 10-year Treasury yields and US market earnings yield of +0.81 throughout the entire 1965-2001 period (Asness, 2003)[49].

Weigand and Irons (2006) argue that a conflict arises between the Campbell and Shiller model and the Fed Model due to the fact that the former assumes the market's price-to-earnings ratio is a stationary mean-reverting time series, while the latter posits that the market's earnings yield (and hence price-to-earnings) is non-stationary. The reasoning that goes behind the Fed Model's assumption of

[48] This Fed Model model was not officially postulated by the US Federal Reserve, but was named as such by market participants as it emerged from data contained in Board of Governors of the Federal Reserve System Monetary Policy Report to the Congress (1997).
[49] Full analysis of the Fed Model is beyond the scope of this paper. For an excellent overview and analysis on the historical relationship behind 10-year yields and the S&P 500's earnings yield, the lack of rigorous theoretical justifications behind the Fed Model, as well as analysis of long-term stock return forecastability of the Fed Model, see Asness (2003).

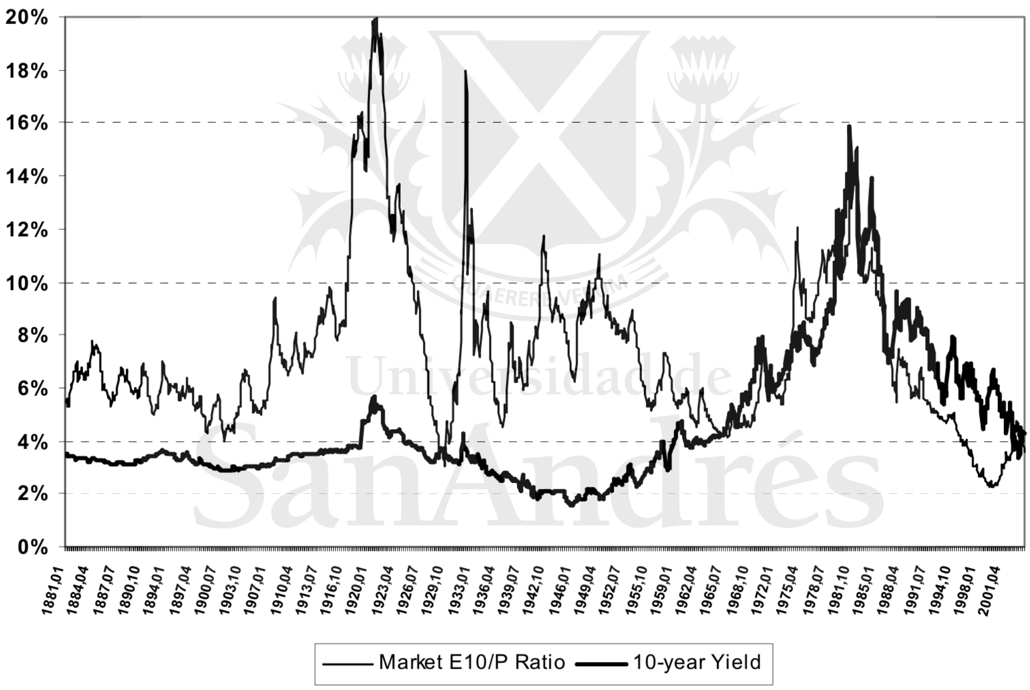
non-stationary price-to-earnings ratios is as follow: given that the market's earnings yield is supposedly tracking the yield on the 10-year Treasury bond, which is itself known to be a non-stationary time series (Bradley and Lumpkin, 1992; Mehra, 1996; Tatom, 2002), it must hold that the market's price-to-earnings ratio is also a non-stationary series. As the authors argue, the Fed Model implications are thus in direct conflict with Campbell and Shiller's model assumption that valuation ratios are stationary and thus mean-reverting.



<u>FIGURE 6</u>: Comparison between the US Stock Market's valuation levels (as represented by the earnings yield) and the yield on the US 10-Year Government Treasury, 1982-1997. Source: US Federal Reserve (1997)

The authors perform augmented Dickey-Fuller unit root tests to 10-year Treasury bond yields, and the market's real price-to-earnings/earnings yield ratios (both 1-year and cyclically-adjusted 10-year) between 1881 and 2004 using monthly data points. They find that for the 1881-1959 sample, the market's price-to-earnings ratio (standard 1-year) and 10-year cyclically-adjusted price-to-earnings ratio (CAPE-10) were stationary throughout the period. Most significant, however, are their results that *all* variables (10-year Treasury yields, the market's 1-year price-to-earnings ratio and 10-year cyclically-adjusted price-to-earnings ratio) are nonstationary (stochastic) from

1960 to 2004, as the correlation between bond yields and equity market earnings yields increased (see Figure 7). Furthermore, they find evidence that the 10-year Treasury bond yield and the market's earnings yield have been cointegrated since 1960, supporting the theoretical underpin behind the Fed Model and validating similar conclusions from Yardeni (2003) and Asness (2003). These findings are consistent with the notion that US investors had been pricing equities such that their real earnings yields were comparable to 10-year nominal yields. Blanchett, Finke, Pfau (2013) find a negative (positive) relationship between S&P500 cyclically-adjusted price-to-earnings ratios (cyclically-adjusted earnings yield) and bond from 1960 to 2012, with an R^2 of 35.6%, offering further supporting evidence of the Fed Model.



<u>FIGURE 7</u>: 10-Year Treasury Yields compared to the market's cyclically-adjusted 10-year earnings yield (reciprocal of CAPE-10), 1881-2004. The high degree of correlation that began in 1960, after which the time series were cointegrated, is similar regardless of whether one uses CAPE-10 or "conventional" 1-year earnings-yield.
Source: Weigand and Irons (2006)

Weigand and Irons note how these results appear to have significant consequences for Campbell and Shiller's model of return predictability on the basis of mean-reverting valuation ratios (cyclically-adjusted ratios or otherwise).

While the authors are quick to admit that the theoretical underpin behind the Fed Model rests on a behavioral error[50] (with little in the way of theoretical underpinnings) and even suggest that the cointegration between 10-year Treasury yields and the market earnings yield could well disappear if interest rates rise substantially from current levels[51] (and thus return to Campbell and Shiller's assumptions of stationary series), the authors make an implicit point that, should investors remain tied to the Fed Model heuristic, valuation ratios could continue to be non-stationary and return predictability using a model that assumes the opposite (i.e.: CAPE model) will likely work less effectively than it has in the past.

The literature that has focused on the question of mean reversion in (non-smoothed) price-to-earnings ratios has been more mixed. Carlson, Pelz and Wohar (2002) show that the empirical "normal" price-to-earnings and price-to-dividend ratios have changed over time. On the other hand, Becker, Lee, Gup (2012), investigating the issue in US markets from 1871 through 2003, find that there is evidence for stationary price-to-earnings ratios, but that this occurs around multiple breaks throughout the period; while Davis, Aliaga-Díaz and Thomas (2012) support the view that price-to-earnings ratios had mean-reverting tendencies in their sample from 1926 through 2011, although they find that the level itself is not constant over time. Campbell and Shiller themselves also examined the possibility of price-to-earnings ratios exhibiting different behavior than they had in the past. When evaluating the poor forecasted long-term returns in the early 2000 given extremely elevated CAPE levels during those times, Campbell and Shiller (2001) mentioned: "We do not find this extreme forecast credible; when the [cyclically-adjusted price-to-earnings ratio] has moved so far from the historically observed range, we cannot trust a linear regression line."

[50] Weigand and Irons (2006) argue that investors are comparing a real variable in the earnings yield versus a nominal variable in the 10-year Treasury bond yield, similar to the argument put forth by Asness (2003).
[51] As of 2015, this is yet to happen, as the 10-year Treasury bond yield has continued to reach successively lower lows since Weigand and Iron's publication.

D. Demographics and Long-Term Stock Market Valuation Ratios

Some authors have suggested that long-term, multi-decade demographic shifts are responsible for driving underlying demand for financial assets as a whole (stocks and bonds). The theory argues that, as individuals enter their high-wage and high-savings years, their marginal propensity to purchase stocks and bonds increases.

While not addressing the CAPE ratio directly, Geanakoplos, Magill and Quinzii (2004) find a statistical link between short-term (1-year trailing) P/E ratios and the *middle-young* (MY) US population cohort ratio (the ratio between the size of the 40-49 year cohort and the 20-29 year cohort) over 20 year horizons, as seen in Figure 8[52]. The authors conclude that *changes*[53] in the MY ratio can be used to predict equity returns (absolute returns and excess returns) in US markets (the international evidence is mixed).

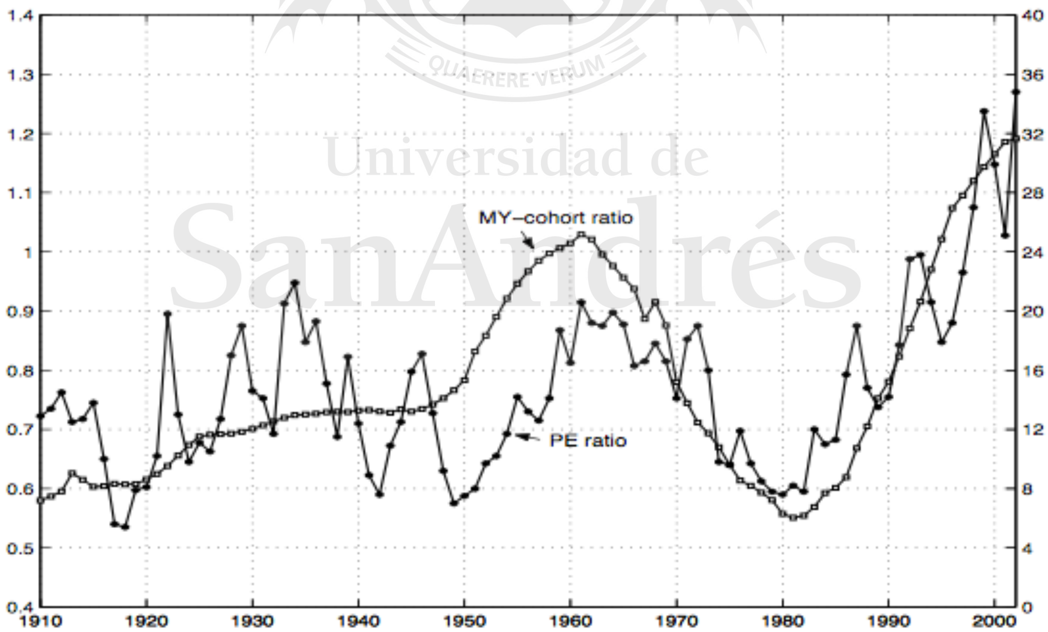


FIGURE 8: S&P500 price-to-earnings ratio and MY cohort ratio. Source: Geanakoplos, Magill and Quinzii (2004)

[52] The evidence is weaker, but still significant, at shorter time frames.
[53] As opposed to *levels* of the MY ratio.

Liu and Spiegel (2011), likewise analyze the relationship between P/E ratios and demographics. Referencing Geanakoplos, Magill and Quinzii (2004), the authors attempt to use the *middle-old* cohort (M/O) (the ratio between the 40-49 and the 60-69 cohorts) to attempt to find similar relationships. In their view, "*the saving and investment behavior of the old-age cohort is more relevant for asset prices than the behavior of young adults*". Consequently, they find a 61% correlation between the M/O cohort ratio and P/E ratios during the 1954 to 2010 sample period, suggesting the overall "pool" of potential stock and bond purchasers help drive valuation ratios.

By themselves, these results pose another threat to the concept of informational efficiency in markets: demographic shifts do not occur in a random manner, and strategies that anticipate the evolution of P/E ratios on this basis could potentially be devised such that participants could benefit not only from long-term real earnings growth but also from additional P/E multiple expansion. Most importantly, while there has been little in the way of studies linking CAPE ratios to demographic cohorts, similar implicit conclusions could be drawn regarding the usefulness of price-to-earnings ratios (smoothed or otherwise) to predict long-term returns: if, as these authors suggest, such price-to-earnings ratios are heavily influenced by the overall availability of buyers (a flow argument), then their predictive ability regarding future returns – and the assumption of mean-reversion in these ratios – should not be examined independently but rather in the context of long-term expected demographic shifts[54].

We would anticipate that these results by themselves may not necessarily discredit Campbell and Shiller's findings. Even if one assumes price-to-earnings changes are fully explained by demographic factors, insofar as these cohort ratios mean-revert over time, cyclically-adjusted price-to-earning ratios should continue to serve as valid tools to reasonably forecast future returns. Campbell and Shiller (1998) criticize the demographic argument as they argue that the expected paths of future earnings and dividends will not be affected by an

[54] Geanakoplos, Magill and Quinzii (2004) and Liu and Spiegel (2011) have emphasized the increasing globalization of equity markets and widening availability of foreign capital as a development that may undermine this relationship in the future.

increase in the current demand for equities; hence, they argue, higher equity prices will merely serve to lower future returns over the long term, consistent with the empirical evidence. It would likely take a considerably more extreme scenario, where demographic cohort ratios are fully responsible for driving equity valuations *and* where said ratios are non-stationary time series, for models that assume otherwise – such as Campbell and Shiller's – to lose their forecasting appeal[55].

[55] For further discussion on the concept of Non-Stationarity in valuation ratios, refer to Section 2.C.

Part III: Beyond the S&P500, Testing the Campbell and Shiller Cyclically-Adjusted Price-to-Earnings Ratio Outside the Traditional Dataset

A. Evidence in Developed and Emerging Markets

A number of authors have attempted to study the predictive power of Campbell and Shiller's cyclically-adjusted price-to-earnings ratio in equity markets beyond the United States, both developed and emerging. Much like the case of the original P/E anomaly – which was discovered using US equity market data and for which research targeting developed and emerging markets surfaced thereafter[56] – the main motivation has been to study the existence of these anomalies on out-of-sample data.

By obtaining historical earnings and price data from equity markets in 44 countries (both developed and emerging)[57] from 1980 to 2014, Faber (2014) constructs 10-year cyclically-adjusted price-to-earning series for each country (as far back as the data permits). The author finds a similar relationship as that described for the S&P500 by Arnott (2011): there is a significant negative relationship between starting CAPE10 ratio and subsequent 10-year real returns (study based in US Dollars), with a delta between top and bottom quintiles of almost 10% annualized. Faber then proposes a trading strategy that systematically invests in the cheapest tercile or quartile countries by CAPE10 ratio and rebalances on an annual basis (importantly, the system evaluates cheapness relative to the investable universe at the time, which increases from around 15 countries in 1980 to the full 44 country sample by the late 1990's). Both cases showed 3-4% of real annual outperformance over an equal-weighted portfolio (composed of equal allocations to all the countries in the entire country

[56] Broad country results available in Fama and French (1998) and Arshanapalli, Coggin, and Doukas (1998), while other country-specific studies include: Brouwer, van der Put and Veld (1997) and Bird and Whitaker (2003), for United Kingdom; Aggarwal, Rao and Hiraki (1990) and Chan, Hamao and Lakonishok (1991) for Japan; and Kelly, McClean and McNamara (2008) for Australia.
[57] Of which 21 countries are classified as emerging markets, and 23 are developed markets.

universe), and despite the slightly higher volatility managed to produce higher risk-adjusted returns. Equally, a strategy of investing in the 25% or 33% most *expensive* countries in the country universe and rebalancing on an annual basis showed roughly 3% of underperformance relative to the equal-weighted country index, with higher volatility and thus materially lower risk-adjusted return profile. The author further addresses the potential criticism of data mining for choosing an arbitrary 10-year holding period when calculating real returns: re-examining results for holding periods of 1, 3, 5 and 7 years showed no material difference to the 10-year case, as all configurations managed to produce higher absolute and risk-adjusted returns throughout the sample period than the buy-and-hold equal-weight approach. Drawdowns were not materially different across the cases, and the author notes that repeating the study in local currency-terms produced no material difference in real terms as the dollar-denominated study.

Similar results are found by Butler, Philbrick, Gordillo and Faber (2012) as they examine real equity returns across 32 countries from 1999 to 2011 for the cheapest 33% of countries on a CAPE10 basis: absolute returns are higher by around 2.5% on an annual returns (relative to an equal-weight portfolio of all countries), but the strategy had sufficiently higher volatility that risk-adjusted returns were only marginally better (Sharpe ratio of 0.44 for the CAPE10 model versus 0.41 for the equal weight basket), and suffered from slightly higher drawdowns as well. The authors propose volatility targeting at the portfolio level[58] and risk parity[59] approaches to reduce the higher volatility and drawdowns associated with global CAPE trading models. The authors find that exploiting the value anomaly by purchasing a basket of the cheapest countries

[58] Scaling exposure accordingly for the volatility of the strategy to match a pre-determined target, in this case 10% annualized.

[59] These strategies do not attempt allocate an equal amount of capital across all markets that fall under the cheapest 33% of countries (as Faber does) but rather allocate weightings such that the contributions to portfolio volatility from each individual market is equal (i.e.: volatile markets will receive lower allocations, all else being equal)

globally and targeting volatility at the portfolio level can substantially improve results relative to both equal-weighted and valuation-driven models alike[60].

Similar work across international markets is performed by Klement (2012), who studies the Campbell and Shiller CAPE10 as a return forecaster in 35 countries, 19 of which are developed and 16 emerging, up to 2012 (the starting date of the CAPE series is country-specific)[61]. The author finds the metric to be a solid predictor of future returns in developed markets: using a panel regression methodology (to avoid autocorrelation issues) yields an R^2 of 33% for the log of initial CAPE10 and subsequent annualized 10-year real return across the entire developed country sample (significant at the 99% confidence level). In the case of emerging markets, the author finds the issue of limited data to be problematic with CAPE10 and thus runs a separate regression reducing the forecasting period to 5 years. This regression provides an R^2 of 18%, higher than all other competing forecasting variables. The author sets out to examine the link between the Campbell and Shiller CAPE and macroeconomic variables across countries, finding an overall positive relationship between GDP growth and market CAPE10 levels and between real interest rates and CAPE10 ratios and a negative relationship between inflation and CAPE10 levels in the sample, in overall agreement with previous findings. Klement (2013) revisits the study and further focuses on relationships between market CAPE10 levels and macroeconomic variables: the author runs regressions across the country universe studying preceding 10-year change in 10-year bond yields (local currency) and current CAPE10 levels and finds statistically strong negative correlation between the variables throughout the country universe. The author further updates the study with another year of returns, and maintains the

[60] Butler, Philbrick and Gordillo's Risk Parity CAPE model produced 12.2%, 1.06, and 31% real annualized returns, Sharpe ratio, and maximum drawdown, respectively. Faber's CAPE model produced 10.8%, 0.44 and 65.1%, and an equal-weight basket produced 8.4%, 0.41 and 58.4%, respectively, during the same time period (1999-2011).

[61] The author notes the difficulty of assembling emerging market data considering not only prior episodes of severe macroeconomic shocks that distort earnings and price series (mass nationalizations, hyperinflation, and so forth), but also the relative short duration of the series. In emerging markets with the least amount of data history, for example, the author notes that the figures begin in 1998 and given the 10-year lag used to compute the CAPE10 metric, return analysis can only be established for about 4 years (2008 – 2012). Despite these country-specific shortcomings, most countries have materially longer available data.

conclusion that, with rare individual exceptions, the CAPE10 is a significant forecaster of future real stock market returns across developed and emerging markets alike[62].

B. Evidence in Individual Stocks

While the Campbell and Shiller cyclically-adjusted price-to-earnings ratio was conceived as a metric that predicted the future return potential of the S&P500 index, and research since has largely focused on its ability to predict returns for the aggregate stock market, a limited area of focus has been applying the CAPE10 methodology on individual stocks. This represents a departure from the traditional cross-sectional value anomaly work previously highlighted[63]: whereas prior studies evaluated a stock's relative valuation on the basis of the most recent 12 months of financial data (earnings, for example), the studies covered in this subsection use CAPE ratios. The rationale is to minimize the earnings volatility at the company level as a result of macroeconomic and other external shocks, such that a smoothed earnings series (throughout the business cycle) might prove to be a more accurate approximation the underlying company's earnings power. Gray and Vogel (2013) set out to study the ability of the CAPE10 and other cyclically-adjusted measures (cyclically-adjusted real book to market, real free cash flow to enterprise value, and others) to forecast returns on all individual US listed stocks from 1973 to 2012.

Using annual rebalancing, the authors found significant outcome differences across starting CAPE10 deciles, with a delta between cheapest and most expensive decile of almost 6% in annualized real returns, and a substantial gap in risk-adjusted metrics (Sharpe ratio of 0.27 for the growth decile vs. 0.63 for

---

[62] This subsection covers studies by market practitioners, as there is understandably very little in the way of academic studies on backtested performance of CAPE models in emerging and developed markets. While said results may be viewed somewhat more skeptically given the lower potential degree of research rigorousness (given the lack of peer reviews and so forth), it bears reminding that authors are broadly in agreement as to the conclusions derived.
[63] For further information, refer to Appendix A.

the value decile) and downside risk metrics (Sortino[64] ratio of 0.40 for the growth decile vs. 0.84 for the value decile). The difference was even more pronounced when using monthly rebalancing (almost 10% annualized real return difference, with higher Sharpe and Sortino ratio gap). The authors found no significant difference in drawdown performance between the cheap and expensive deciles. In both cases, comparing the results against an equal-weighted basket composed of all the stocks in the investable universe provided a clear advantage to the CAPE10 model: up to 5.8% difference in annualized returns, with higher Sharpe and Sortino ratios.

The drawdown profile of the CAPE10 models versus the equal-weight portfolio was nonetheless weaker, similar to Gordillo, Philbrick and Butler (2012) results using global CAPE trading models. In line with those authors, Gray and Vogel attempt to expand on the simple CAPE strategy by implementing a momentum overlay, finding it improves risk-adjusted returns further and minimizes drawdowns relative to the base strategy. Interestingly, the author's proposed Cyclically-Adjusted Book-to-Market ratio (following preference for the original metric in Fama and French, 1992) yields the strongest performance of all strategies in absolute and risk-adjusted terms.

C. Evidence Across S&P500 Sectors

Ural, Lazanas, Zhuang and Staal (2012) study the relationship between CAPE-10 and returns on a sector basis[65] for S&P500 firms over the December 1982 to May 2012 timeframe. Using a new dataset consisting of S&P500 sector earnings and return data over 40 years, the authors initially find strong negative correlations between starting CAPE ratio and subsequent total returns over

[64] The Sortino ratio is similar to the Sharpe ratio, in that it attempts to examine excess returns above the risk-free rate per unit of risk, but it does so by utilizing downside volatility (below a minimum acceptable return threshold) instead of pure volatility.

[65] As described by the Global Industry Classification Standard (GICS), representing: Energy, Materials, Industrials, Consumer Discretionary, Consumer Staples, Health Care, Financials, Information Technology, Telecommunication Services and Utilities.

multiple timeframes on a sector basis (in line with previous findings on an aggregate index level).

The authors then attempt to build a systematic trading system using the CAPE methodology: the strategy consistently invests in an equal-weighted basket of the cheapest 50% (5 out of 10) of S&P500 sectors – as represented by the CAPE-10 – and rebalances on a monthly basis. To correctly interpret the relative attractiveness of one sector versus another, they introduce the *Relative CAPE Indicator*, which is merely the ratio of the current CAPE-10 value for a sector relative to its long-term average[66] (all else equal, a higher value is less attractive and makes the sector less likely to be included for investment for that particular month)[67]. The system shows significant outperformance in absolute and risk-adjusted terms, with lower maximum drawdowns versus a passive S&P500 investment strategy. Just as importantly, an even more significant performance difference is found when comparing against a strategy that systematically invests in the 50% most *expensive* S&P500 sectors (also on a relative CAPE indicator basis). The authors then attempt to take the model further by including a momentum filter (of the 5 cheapest sectors, excluding the one with the worst trailing 12-month total return): they find the value and momentum combination produced the best absolute and risk-adjusted returns throughout the 1982 to 2012 timeframe, and with an extremely high degree of consistency on a year-to-year basis.

Similar work is performed by Bunn and Shiller (2014), who apply the new CAPE ratio on a sector basis from 1870 to 2012 and, in line with prior literature, find significant evidence of return predictability on 10-year timeframes given the initial CAPE ratio. They also build upon the Relative CAPE Indicator introduced by Ural, Lazanas, Zhuang and Staal (2012) and devise a trading model that overweights the single cheapest S&P500 sector on a Relative CAPE basis and underweights the single most expensive one, finding over 1% of annualized

[66] 20-year average, windsorised at 5% level to remove outliers.

[67] The authors work alongside Bunn and Shiller in developing the Relative CAPE Indicator and implementing methodology changes in the CAPE ratio, which is detailed in Bunn and Shiller (2014)

inflation-adjusted excess returns from the strategy throughout the 1870 to 2012 period. Bunn and Shiller summarize: "Our results using over a hundred years of data are consistent with the notion that major sectors of the stock market show frequent mispricings that can be exploited in an investment strategy that generally leads to better results than holding the market portfolio."

The findings of both papers, as applied to a new dataset composed of granular sector-level prices and earnings, validate the usefulness of the CAPE ratio in predicting not only absolute returns on a sector basis, but also in predicting relative returns across sectors given differences in starting Relative CAPE levels.

Part IV: Conclusions

Campbell and Shiller's Cyclically-Adjusted Price-to-Earnings ratio has once again come into the spotlight, as current levels for the S&P500's CAPE10 continue to predict substantially poor long-term returns. While Bunn and Shiller (2014) have implemented methodology changes to account for new capital market developments that had been a particular point of argument with other researchers (the most significant change was to capture the ever-increasing importance of share buybacks), the debate continues to largely center around the issue of risk.

The main argument against valuation ratios as predictors of subsequent stock returns has traditionally focused on the concept of investment risk, and similar arguments can be levied against Campbell and Shiller's metric: these authors suggest that if present valuation metrics correspond to higher-than-usual expected returns, it must be due to higher-than-usual levels of risk, and vice-versa. The body of empirical evidence as covered in this paper has not seen such a relationship on an ex-post basis: lower-than-average CAPE10 levels have in fact been shown to correlate with *superior* excess returns (and vice-versa) in a risk-adjusted framework. This is regardless of the many "rules" that such a system can opt to take[68] and of the definition of risk: whether investment risk corresponds to return volatility, maximum drawdowns, downside volatility, or a combination thereof. Just as significantly, the CAPE10 has been shown to predict subsequent returns in out-of-sample data: data for S&P500 returns, prices and earnings generated since the model's formulation (the 1988 to 2014 period); emerging markets returns; individual stocks; and across S&P500 sectors.

Nonetheless, much like Fama and French's (1992) conclusion that excess returns from small and value firms correspond to bearing unidentifiable risks, it is hard to deny that historically low CAPE10 readings have tended to coincide with periods of extreme perceived risk and stress in the financial system: the episodes

[68] In terms of the rules used to dictate the level of cash and equity allocation depending on existing CAPE10 levels.

include economic depressions (1921, 1933), oil shocks (1974) and runaway inflation (1982). In these instances, the abnormally-high realized returns from holding equities is naturally unsurprising; in the context of risk-adjusted returns and to the detriment of the EMH view, however, the empirical evidence has been shown to be conclusive in such cases, as equity holders have historically been compensated above and beyond their assumed level of investment risk. Therefore, in light of the evidence, it must hold that either markets were compensating investors for assuming unidentified risks or the apparent anomaly owes itself to behavioral considerations. On the former point, the possibility remains that unidentified, unquantifiable risks were well present within those times such that the market was properly pricing-in the catastrophic "left tail" scenario: the absence of such an outcome in US market history does not negate its existence in the distribution of outcomes.

Barring such conclusions, which by definition are unworkable from a quantitative standpoint, one could naturally gravitate to the second argument: that the CAPE anomaly might be caused by investor's tendency to vary their risk aversion levels in response to the macroeconomic environment due to behavioral reasons (amongst which we can highlight the extrapolation of current trends, overconfidence in growth expectations and herding behavior). This is supported by empirical findings of an optimal combination of macro variables that results in systematically higher (cyclically-adjusted) earnings multiples[69]. In such a framework, positive risk-adjusted returns from investing in low starting CAPE periods may be compensation for bearing psychological risks by defying behavioral biases and taking a contrarian approach.

The key contribution in Campbell and Shiller's work seems to be the process through which ratios are smoothed to account for influences arising from business cycle changes and external shocks, and not in the choice of the actual smoothing time period (10 years in the original formulation), forecast horizon (10 years originally) nor of the actual time series used (index earnings per share originally). Indeed, authors such as Gray and Vogel (2013) have shown that

[69] Parikh (2012) and Arnott (2011). Refer to Section 1.H.

other cyclically-adjusted valuation metrics (book-to-market, for instance) had as high or higher explanatory power of long-term equity returns as Campbell and Shiller's CAPE10. Notwithstanding its potential shortcomings, that is the assumption of mean-reversion in price-to-earnings ratios (which continues to be the subject of much debate) and ever-changing accounting and inflation calculations (that have opened valid criticisms regarding historical comparability may require methodological changes to keep up to date)[70], Campbell and Shiller's CAPE10 should continue to be one of the preferred tools for evaluating prospective long-term stock market returns within a risk-adjusted framework.

We suggest further research on the topic should continue to focus on: gathering new evidence in emerging markets, as new data is generated on an ongoing basis and its reliability is enhanced; study the direct link between CAPE ratios and demographic cohorts; extend Gray and Vogel's (2013) contribution and analyze the performance of varied cyclically-adjusted ratios (variations of price-to-book, price-to-dividends and price-to-cash flow) on *aggregate* stock market performance, using these new metrics both individually and as a composite; and further examine the integration of value (as represented by CAPE10) and momentum in a risk-adjusted framework given their tendency to be negatively correlated over time (Asness, Moskowitz and Pedersen, 2013) and particularly during extreme market conditions (Moskowitz, Ooi, and Pedersen, 2012).

[70] Naturally, one could highlight the potential for the market pricing-in these unidentified risks to be behind the entire anomaly and hence invalidate Campbell and Shiller's conclusions. Given the natural uncertainty regarding this case, the concrete weaknesses in the model are highlighted instead.

Appendix A: Background on Return Anomalies

There has been a host of academic research investigating the predictability of stock returns on the basis of (trailing[71]) valuation metrics. Most studies have tended to focus on four of the most significant and well-known financial ratios: earnings-to-price (E/P), dividends-to-price (D/P), book-to-market (B/M) and cash flow-to-price (C/P) (or their respective reciprocals: price-to-earnings, price-to-dividend, price-to-book and price-to-cash flow ratios), though the research methodology itself has varied significantly over time. We can both find academic studies that analyzed the relationship between valuation ratios and future stock market performance applied at an index or country level, and those that performed similar studies on a panel data comprised of subset of stocks over a given period.

One of the earliest papers researching the relationship between valuations (trailing one-year price-to-earnings ratios) and subsequent stock performance at an individual company level was performed by Nicholson (1960): in a study involving 100 common stocks in the US market over the 1939 to 1959 period, Nicholson found evidence for a strong inverse relationship between "cheapness" (according to P/E quintile rankings) and total returns across 5, 10, 15 and 20-year timeframes. In a follow-up paper, Nicholson (1968) extended his analysis to include 189 companies through the 1937 to 1962 periods, again finding consistent outperformance of average of stocks in the cheapest quintiles. These results were consistent regardless of the valuation metric used: similar results were obtained while ranking based on price-to-earnings (P/E), price-to-book (P/B) or price-to-sales (P/S).

Further work examining the relationship between P/E ratios and subsequent performance of equities was done by Basu (1977): studying 1,400 US-listed securities from 1956 to 1971, the author concluded low P/E stocks produced

[71] By "trailing" we refer to ratios constructed by applying the last 12 month's worth of financial data to current prices. For example, the trailing P/E divides the current stock price by the last 4 quarter's worth of earnings per share (EPS).

significant excess returns vs. the highest P/E buckets. Most importantly, the same conclusions were reached when examining absolute and risk-adjusted returns, the latter conclusion derived from the fact that there was no evidence of a relationship between average beta (degree of systematic risk incurred) values across the different P/E quintiles. This was in apparent violation of the semi-strong form version of the Efficient Market Hypothesis (Fama, 1970)[72]. Ball (1978) attempted to justify these seeming violations of the Efficient Market Hypothesis by claiming prior studies exhibited several flaws: ignoring the impact of transaction costs, experimental errors and, most importantly, the lack of inclusion of certain risk factors that, if present, claimed would eliminate the apparent P/E anomaly.

The relationship between price-to-earnings ratios and returns is re-examined by Basu (1982) following criticisms by Banz (1981) and Reinganum (1981), which argued for firm size (and not price-to-earnings) as the primary determinant of subsequent risk-adjusted returns. These authors claimed that the size factor "subsumes" the P/E effect and that, much like Ball (1978), both are mere proxies for other undiscovered factors. Basu finds that the earnings yield effect on subsequent absolute and risk-adjusted returns is significant even after accounting for firm size, but that the significance of the earnings yield effect was inversely related with firm size (i.e.: higher significance for lower firm size and vice versa). Fuller, Huberts and Levinson (1993) also studied the assertions of Ball (1978) – who claimed that the P/E effect was merely a proxy for another underlying undiscovered factor – by analyzing the performance of about 1,000 of the largest US stocks between 1973 and 1991 using a multi-factor model that not only incorporated Sharpe's systematic risk (beta) but also added 13 other factors, such as financial leverage, liquidity and earnings volatility, as well as controlling for 55 separate industry classifications. They found that the P/E factor explained most of the difference in excess returns across portfolios, with

[72] Fama (1970) postulated that the so-called "semi-strong" form version of the Efficient Markets Hypothesis implies that asset prices incorporate all available public information, such that excess returns cannot be systematically harvested by the use of fundamental information such as valuation ratios.

the price-to-book being the second most important variable. Significantly, even though subsequent earnings growth over the next 4 years was positively correlated with starting P/E – indicating investors were pricing stocks correctly on the basis of future growth prospects – they found that earnings growth did not prove useful in forecasting excess returns.

A potential explanation for these apparent size and P/E anomalies was put forth by Banz and Breen (1986), who suggested that the data sources on which many of these studies were based suffered from two major biases: ex-post selection bias, meaning the exclusion of companies that had previously disappeared (via bankruptcy, mergers, or other similar corporate events); and look-ahead bias, implying that portfolios were built on the basis of accounting information that would not have been available at that exact moment (given the typical lag between the end of fiscal reporting periods and subsequent public dissemination of said information). The authors found that the P/E effect was not significant after correcting for these biases in the underlying data; though the size effect was still present.

Further work on the the effects of size and value factors (the latter represented by book-to-market ratio) was performed by Fama and French (1992): aiming to test the validity of the prevailing one-factor CAPM (systematic risk) by studying stock returns during the 1962 to 1989 period in the US market, the authors found that the size and value factors offer significant explanatory evidence on the cross-sectional dispersion of security returns. The authors suggest that the excess returns earned by the small/value firms might be merely rewards for bearing other unidentifiable risks, given the evidence for low book-to-market firms to offer "persistently weak" economic performance (and vice versa).

While most research up to this point had focused on the concept of investment risk when analyzing prior findings, either explicitly via beta (systematic risk) or implicitly through factors such as size, DeBondt and Thaler (1985) examined the problem from a behavioral standpoint: they proposed that the excess returns from low P/E stocks are derived from an initial overreaction to apparent negative news, triggered by investor failure to apply Bayes' rule when acting upon newly obtained information (Kahneman and Tversky, 1971), such that they become overly pessimistic and price securities in such a manner that obtaining positive excess risk-adjusted returns in such securities is feasible from that point onwards (the opposite holds true for high P/E stocks). Upon testing the predictive effect of the overreaction hypothesis (purely on the basis of historical return data and excluding any accounting-related data such as earnings), they found significant evidence in favor of this apparent anomaly in US-listed common stocks between 1926 and 1982: portfolios composed of prior "loser" stocks (trailing 36 month timeframe) outperformed their prior "winner" counterparts by 25% on average over 3-years after portfolio formation, and did so with significantly lower risk, thus earning excess risk-adjusted returns. This held even after controlling for firm size, dividend yield and financial leverage.

Similar "contrarian" investment strategies were studied by Lakonishok, Schleifer, and Vishny (1994): the authors divided firms across "value" and "glamour" subsets on the basis of valuation metrics such as cash flow-to-price, book-to-market, earnings-to-price and historic sales growth, finding the "value" firms outperform their "glamour" counterparts with no higher systematic risk, even after controlling for firm size and leading to the conclusion that investors potentially overly-extrapolate past growth when pricing securities. Dreman and Lufkin (1997) found similar evidence for the existence of both the value and firm size factors in US stocks between 1970 and 1995.

Just as importantly, the existence of the so-called value anomaly has been verified in international markets: in the United Kingdom, by Brouwer, van der Put and Veld (1997) and Bird and Whitaker (2003); in Japan by Aggarwal, Rao and Hiraki (1990) and Chan, Hamao and Lakonishok (1991); in Australia by Kelly, McClean and McNamara (2008); in multiple international markets by Fama and French (1998) and Doukas, Arshanapalli and Coggin (1998). For a more complete list on the literature of the P/E effect and size effects in worldwide markets, refer to Khorsand and Ahmed (2014) and Pathirawasam (2010), respectively.

It is worth mentioning further evidence put forth for the existence of other empirical return anomalies: Bhandari (1981) found a positive relationship between stock returns and company leverage; Stattman (1980); Rosenberg, Reid, and Lanstein (1984) and Chan, Hamao, and Lakonishok (1991) found that the book-to-market ratio was able to better explain returns than the traditional one-factor CAPM model; Jegadeesh and Titman (1993) found evidence in favor of momentum, under which recent return performance correlated strongly with future performance over 3-12 month horizons; Daniel and Titman (2006) find a negative relationship between the amount of stock issuance by firms and average returns (companies that fund their growth via high levels of stock issuance have poor future returns and vice-versa). Further background on the uncovered array of return anomalies in the literature, refer to Beechey, Gruen and Vickery (2000) and Fama and French (2007).

Bibliography

- Aggarwal, R., Rao, R. P., and Hiraki, T., 1990, "Regularities In Tokyo Stock Exchange Security Returns: P/E, Size, And Seasonal Influences". Journal of Financial Research, Southern Finance Association; vol. 13(3), p. 249-263.
- Angelini, N., Bormetti, G., Marmi, S., Nardini, F., 2012, "Value Matters: Predictability of Stock Index Returns"
- Arnott, R., 2011, "King of the Mountain", Research Affiliates, available at: http://www.researchaffiliates.com/Production%20content%20library/F_2011_Sept_King_of_the_Mountain.pdf
- Asch, S. E., 1952, "Effects of group pressure on the modification and distortion of judgements". In G. E. Swanson, T. M. Newcomb & E. L. Hartley (Eds.), *Readings in social psychology* (2nd ed., pp. 2–11).
- Asness, C. S., 2002, "Fight the Fed Model: The Relationship Between Stock Market Yields, Bond Market Yields, and Future Returns", AQR Capital Management.
- Asness, C. S., 2012, "An Old Friend: The Stock Market's Shiller P/E", AQR Capital Management.
- Asness, C. S., Moskowitz, T. J., and Pedersen, L. H., 2013, "Value and Momentum Everywhere", *Journal of Finance*, 68 (3), 929-985.
- Ball, R., 1978, "Anomalies in Relationships Between Securities' Yields and Yield surrogates", Journal of Financial Economics 6, 103-126.
- Barberis, N., Schleifer, A., and Vishny, R., 1998, "A Model of Investor Sentiment", Journal of Financial Economics 49: 307-343.
- Basu, S., 1977, "Investment Performance of Common Stocks in Relation to their Price-Earnings Ratios: A Test of the Efficient Markets Hypothesis", *Journal of Finance*, 32:3, 663-82.
- Basu, S., 1982, "The Relationship Between Earnings' Yield, Market Value and Return for NYSE Common Stocks", *Journal of Financial Economics*, 129-56.
- Beechey, M., Gruen, D., and Vickery, J., 2000, "The Efficient Market Hypothesis: A Survey", Economic Research Department, Reserve Bank of Australia.
- Becker, R., Lee, J. and Gup, B. E., 2012, "An empirical analysis of mean reversion of the S&P 500's P/E ratios", Journal of Economics and Finance, 36(3): 675-690.
- Bikhchandani, S., Hirshleifer, D., and Welch, I., 1992, "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades", Journal of Political Economy, Vol. 100, No. 5; pp. 992-1026.
- Bird R., and Whitaker J., 2003, "The performance of value and momentum investment portfolio: recent experience in the major European markets", Journal of Management; vol.4, PP.221-246.
- Blanchett, D., Finke, M. S. and Pfau, W. D., 2013, "Optimal Portfolios for the Long Run", SSRN.
- Bradley, M. G., Lumpkin, S. A., 1992, "The Treasury Yield Curve as a Cointegrated System", The Journal of Financial and Quantitative Analysis, Vol. 27, No. 3; pp. 449-463.

- Braun, T., and Kaussen, M., 2014, "Don't Try to Time the Market with the Shiller P/E", BWM Asset Management.
- Brodeski, B., Beall, G., Larson, A., 2012, "Tactical Think Tank: A Fundamental Answer for Tactical Asset Allocation", Journal of Financial Planning.
- Brouwer, I., Van Der Put, J., Veld, C., 1997, "Contrarian Investment Strategies in a European Context", Journal of Business Finance & Accounting, 24: 1353–1366.
- Bunn, O., and Shiller, R., 2014, "Changing Times, Changing Values: A Historical Analysis of Sectors within the US Stock Market 1872-2013", Cowles Foundation for Research at Yale University.
- Butler, A., Philbrick, M., 2011,"Estimating Future Stock Market Returns", Darwin Funds.
- Butler, A., Philbrick, M., Gordillo, R., and Faber, M., 2012, "Global CAPE Model Optimization", Darwin Funds.
- Campbell, J. Y., and R. J. Shiller, 1988, "Stock Prices, Earnings and Expected Dividends,", Journal of Finance, 43 (3), 661-76.
- Campbell, J. Y., and R. J. Shiller, 1998, "Valuation Ratios and the Long-Run Stock Market Outlook," Journal of Portfolio Management, 24:2, 11-26.
- Campbell, J. Y., and Yogo M., 2006, "Efficient tests of stock return predictability", Journal of Financial Economics 81(1): 27-60.
- Campbell, J. Y., and Thompson, S. B., 2008, "Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average?", The Review of Financial Studies 21(4): 1509-1531.
- Carlson, J. B., Pelz, E. A., and Wohar M. E., 2002, "Will Valuation Ratios Revert to Historical Means?", The Journal of Portfolio Management, Vol. 28, No. 4: pp. 23-35.
- Chan, L. K. C., Hamao, Y. and Lakonishok, J., 1991, "Fundamentals and Stock Returns in Japan", The Journal of Finance, 46: 1739–1764.
- Daniel, K. and Titman, S., 2006, "Market Reactions to Tangible and Intangible Information", The Journal of Finance, 61: 1605–1643.
- Davis, J. H., Aliaga-Díaz R., Thomas, C. J., 2012, "Forecasting Stock Returns: What Signals Matter, and What Do They Say Now?", The Vanguard Group.
- De Bondt, W.F.M, Thaler R., 1985, "Does the Stock Market Overreact?"
- Doukas, J. A., Arshanapalli, B., and Coggin, T. D., 1998, "Multifactor Asset Pricing Analysis of International Value Investment Strategies", SSRN.
- Dreman D. N. and Berry M. A., 1995, "Overreaction, Underreaction, and the Low-P/E Effect", Financial Analysts Journal 51 (4): 21–30.
- Faber, M., 2012, "Global Value: Building Trading Models with the 10-Year CAPE", Cambria Quantitative Research.
- Fama, E., 1970, "Efficient Capital Markets: A Review of Theory and Empirical Work". Journal of Finance 25 (2): 383–417.
- Fama, E. F., and French, K. R., 1992, "The Cross-Section of Expected Stock Returns." The Journal of Finance 47: 427–465.
- Fama, E. F., and French, K. R., 1998, "Value Versus Growth: The International Evidence" The Journal of Finance, 53: 1975–1999.
- Fama, E. F., and French K. R., 2007, "Dissecting Anomalies." The Journal of Finance 63, 1653–1678.

- Fisher, K. L. and Statman, M., 2006, "Market Timing in Regressions and Reality". Journal of Financial Research, 29: 293–30.
- Geanakoplos, J., Magill, M., and Quinzii, M., 2004, "Demography and the Long-Run Predictability of the Stock Market." Brookings Papers on Economic Activity 1: 241–307.
- Goddard, K. C., 2011, "Subsuming the Efficient Market Hypothesis", Advisor Perspectives Paper.
- Goyal, A. and Welch, I., 2006, "A Comprehensive Look at the Empirical Performance of Equity Premium Prediction", Yale ICF Working Paper No. 04-11.
- Graham, B., and Dodd, D. L., 1934, "Security Analysis", 1st ed., McGraw-Hill.
- Gray, W. R., Vogel, J., 2013, "On the Performance of Cyclically Adjusted Valuation measures", SSRN.
- Gray, W. R., Wang, T., Zhang, S., Kanner, C., 2013, "Tactical Asset Allocation During Cheap Markets", Empiritrage Research
- Helwege, J., Laster, D., and Cole, K., 1995, "Stock Market Valuation Indicators: Is This Time Different?", Federal Reserve Bank of New York Resarch Paper No. 9520.
- Jegadeesh, N., and Titman, S., 1993, "Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency", Journal of Finance, Vol. 48(1): 65-91.
- Kahneman, D., and Tversky, A., 1971, "Judgment under uncertainty: Heuristics and Biases", Cambridge University Press.
- Kelly, S., McClean, J., and McNamara, R., 2008, "The Low P/E Effect and Abnormal Returns for Australian Industrial Firms", 21st Australasian Finance and Banking Conference, SSRN.
- Khorsand, B., and Ahmed, S., 2014, "Time-Varying Idiosyncratic Volatility, Inter-listing and Value Premium: Evidence from Canadian Market", 21st Global Financial Conference.
- Klement, J., 2012, "Does the Shiller-PE work in emerging markets?", Wellershoff & Partners Ltd., SSRN.
- Klement, J., 2013, "What the Shiller PE Says about Global Equity Markets", Wellershoff & Partners Ltd., SSRN.
- Lakonishok, J., Shleifer, A., and Vishny. R.W., 1994, "Contrarian Investment, Extrapolation, and Risk". The Journal of Finance 49 (5), 1541–78.
- Langer, E., 1975, "The Illusion of Control," Journal of Personality and Social Psychology, 32; 311-328.
- Lewellen, J., 2004, "Predicting returns with financial ratios", Journal of Financial Economics 74 (2004) 209–235
- Liu, Z., and Spiegel, M., "Boomer Retirement: Headwinds for U.S. Equity Markets?", Federal Reserve Bank of San Francisco, 2011
- Malkiel, B. G, 2011, "A Random Walk Down Wall Street: The Time-Tested Strategy for Successful Investing." (Revised Edition)
- Mankiw, N.G., Shapiro, M., 1986, "Do we reject too often: small sample properties of tests of rational expectations models", Economics Letters 20: 139-145.
- Mehra, Y. P., 1996, "Monetary Policy and Long-Term Interest Rates", FRB Richmond Economic Quarterly, vol. 82, no.3; pp. 27-49.

- Montier, J., 2014, “A CAPE Crusader – A Defence Against the Dark Arts”, GMO White Paper.
- Moskowitz, T. J., Ooi, Y. H., and Pedersen, L. H., 2012, “Time Series Momentum”, Journal of Financial Economics 104: 228-250.
- Nelson, C.R., Kim, M.J., 1993, “Predictable stock returns: the role of small sample bias”, Journal of Finance 48, 641-661.
- Nicholson, S. F., 1968, “Price Ratios in Relation to Investment Results”, Financial Analysts Journal (January-February 1968)
- Parikh, S., 2012, “Forecasting Equity Returns in the New Normal”, PIMCO Asset Allocation Focus.
- Pfau, Wade D., Revisiting the Fisher and Statman Study on Market Timing (March 1, 2011). Available at SSRN.
- Roll, R., 2002, "Rational infinitely lived asset prices must be non-stationary," Journal of Banking & Finance, vol. 26(6): 1093-1097.
- Sheikh, A. Z., Qiao, H., 2009, “Non-normality of Market Returns”, JP Morgan Asset Management.
- Shiller, R. J., 1981, “Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends?”, The American Economic Review, Vol. 71, No. 3; pp. 421-436.
- Shiller, R. J., 1996, “Price Earnings Ratios as Forecasters of Returns: The Stock Market Outlook in 1996”, available at: http://www.econ.yale.edu/~shiller/data/peratio.html (posted 21/07/1996).
- Shiller, R. J., 2000, “Irrational Exuberance”, Princeton University Press.
- Siegel, J., 2013, “The Shiller CAPE Ratio: A New Look”, Working Paper.
- Solow, K., Kitces, M. and Locatelli, S., 2011, “Improving Risk-Adjusted Returns Using Market-Valuation-Based Tactical Asset Allocation Strategies”, *Journal of Financial Planning.*
- Stambaugh, R.F., 1986, “Bias in regressions with lagged stochastic regressors”, Working Paper, University of Chicago.
- Stambaugh, R.F., 1999, “Predictive Regressions”, Journal of Financial Economics 54; pp. 375-421.
- Tatom, J. A., 2002, “Stock Prices, Inflation and Monetary Policy,” Business Economics, October, 7-19.
- Torous W., Valkanov R., and Yan S., 2004, "On Predicting Stock Returns with Nearly Integrated Explanatory Variables," The Journal of Business, vol. 77(4), 937-966.
- Ural, C., Lazanas, A., Zhuang, J. and Staal, A., 2012, “Sector Selection Based on the Cyclically Adjusted Price-Earnings (CAPE) Ratio”, Barclays Research.
- Wilcox, S., 2011, “A Cautionary Note About Robert Shiller’s CAPE”, American Association of Individual Investors.
- Weigand, R.A and Irons, R., 2006, “Does the Market P/E Ratio Revert Back to ‘Average?”, Investment Management and Financial Innovations, Volume 3, Issue 3.