

Value Matters: Predictability of Stock Index Returns

Natascia Angelini^{*1}, Giacomo Bormetti^{†2}, Stefano Marmi^{‡3}, and Franco Nardini^{§4}

¹*MEDALics, Dante Alighieri University*

Via del Torrione 95, 89125 Reggio Calabria, Italy

School of Economics, Management and Statistics, University of Bologna

Via Angherà 22, 40127 Rimini, Italy

²*Scuola Normale Superiore*

Piazza dei Cavalieri 7, 56126 Pisa, Italy

³*Scuola Normale Superiore*

Piazza dei Cavalieri 7, 56126 Pisa, Italy

⁴*Department of Mathematics, University of Bologna,*

Viale Filopanti 5, 40126 Bologna, Italy

Abstract

We present a simple dynamical model of stock index returns which is grounded on the ability of the Cyclically Adjusted Price Earning (CAPE) valuation ratio devised by Robert Shiller to predict long-horizon performances of the market. More precisely, we discuss a discrete time dynamics in which the return growth depends on three components: i) a momentum component, naturally justified in terms of agents' belief that expected returns are higher in bullish markets than in bearish ones, ii) a fundamental component proportional to the logarithmic CAPE at time zero. The initial value of the ratio determines

*Electronic address: natascia.angelini@unibo.it

†Electronic address: giacomo.bormetti@sns.it

‡Electronic address: stefano.marmi@sns.it

§Electronic address: franco.nardini@unibo.it; Corresponding author, mobile phone: +39 335 8024631

the reference growth level, from which the actual stock price may deviate as an effect of random external disturbances, and iii) a driving component which ensures the diffusive behaviour of stock prices. Under these assumptions, we prove that for a sufficiently large horizon the expected rate of return and the expected gross return are linear in the initial logarithmic CAPE, and their variance goes to zero with a rate of convergence consistent with the diffusive behaviour. Eventually this means that the momentum component may generate bubbles and crashes in the short and medium run, nevertheless the valuation ratio remains a good reference point of future long-run returns.

JEL codes: G12, G17

Keywords: Valuation Ratios, Long Run Stock Market Returns

1 Introduction

A question that has been asked with a certain regularity since the establishment of the first stock index is the following, see Myers and Swaminathan (1999): is it possible to calculate “the Intrinsic Value of the Dow”? Or, more generally: what is the intrinsic value of Wall Street? A first commonsense answer is that this value may be calculated by summing “the intrinsic value” of each quoted company. At the beginning of the past century, see Williams (1938), Graham and Meredith (1998), it was well known that the latter depends on the expected future earnings and on the balance sheet, but unfortunately there was no easy formula, which yielded even a rough approximation. Therefore aggregating all expectations for all companies quoted in the market soon appeared to be a challenging task.

Between the Fifties and the Sixties the researchers’ attitude towards the problem changed radically. Modern portfolio theory and the statistical evidence that stock prices follow random walks, see Mandelbrot (1965), Fama (1965), has led to the Efficient Market Hypothesis, see Samuelson (1965), Fama (1970). According to this theory, financial markets are “informationally efficient”, that is, one can not achieve returns in excess of average market returns on a risk-adjusted basis, given the information available at the time the investment is made. If this were the case “the intrinsic value” of a stock would be simply its market price and the problem would ultimately be non existent. The bubble burst in 1987 and the exceptional price boom in the late '90 lead more and more researcher to cast doubts on the hypothesis’ truth.

At the very beginning of 2000 Robert Shiller wrote: “we do not know whether the market level makes any sense, or whether they are indeed the result of some human tendency that might be called irrational exuberance”, Shiller (2000), which is apparently at odds with the assumption that market prices are always fair prices. Shiller’s skeptical attitude was anticipated in one of the most famous books ever written on the stock market: in 1934, Graham and Dodd strongly advocated the fundamental approach to investment valuation and recommended to “shift(s) the original point of departure, or basis of computation, from the current earnings to the average earnings, which should cover a period of not less than five years, and preferably seven to ten years” Graham et al. (1962). Since 1988 Campbell and Shiller have been focusing their attention on average earnings as predictor of future dividends Campbell and Shiller (1988a) ¹ and future stock prices Campbell and Shiller (1998) ²; Shiller (2000) proposed an innovative test of the appropriateness of prices in the stock market: the Cyclically Adjusted Price Earning ratio, and showed that it is a powerful predictor of future long run performances of the market; the performance of the test is quite satisfactory in the case of the US market from the end of 19th century up to today. In the same years, again elaborating on Graham et al. (1962), Lander et al. (1997) suggested an alternative but similar criterion based on the assumption “that many investors are constantly making a choice between stock and bond purchases; as the yield on bonds advances, they would be expected to demand a correspondingly higher return on stocks, and conversely as bond yields decline Graham et al. (1962, pag. 510)”; their “approach considers whether stocks are appropriately valued relative to analysts’ perceptions of future earnings and yields on alternative investments” such as high grade corporate bonds (for a comparison between the two approaches see Weigand and Irons (2006)).

The aim of this paper is to provide a theoretical framework which explains how future stock index returns depend on the CAPE ratio. This is achieved by constructing a model. What we require of our model is similar to what Chiarella required in Chiarella (1992, see pag. 102)

“Firstly we require that the model generate a significant transitory component around the equilibrium which reflects the rationally expected value of the asset. Secondly the model must

¹“Our results indicate that a long moving average of real earnings helps to forecast future real dividends. The ratio of this earnings variable to the current stock price is a powerful predictor of the return on stock, particularly when the return is measured over several years”.

²The price-smoothed earnings ratio has little ability to predict future growth in smoothed earnings; the R squared statistics are less than 4% over one year and over ten years. The ratio is a good forecaster of ten-year growth in stock prices, with an R squared statistic of 37%.

allow for the incorporation of chartists, a group which bases its market actions on an analysis of past trends. Since this group seems to be an important part of real markets it is important to determine what effect its activity has on the behaviour of asset prices and whether the behaviour of a model incorporating chartists comes closer to explaining some of the empirical results cited earlier”.

We choose, however, a different approach and we work at an aggregate level instead of explicitly modelling two different types of agents. We introduce a momentum component in a “simple error-correction model that predicts the return of the S&P based on the deviations from a presumed equilibrium between . . . earnings yields” Lander et al. (1997, see pag. 3) and a long run target yield rate. The return growth depends on three components

- a) a momentum component, naturally justified in terms of agents’ expectation that expected returns are higher in bullish markets than in bearish ones;
- b) a fundamental component, proportional to a function of the level of the logarithmic averaged earnings-to-price ratio (for brevity log EP ratio) at time zero; the initial time value of log EP ratio determines the reference growth level, from which the actual stock price may deviate as an effect of random external disturbances;
- c) a driving component ensuring the diffusive behaviour of stock prices.

Under these assumptions, we are able to prove that, if we consider a sufficiently large number of periods, the expected rate of return and the expected gross return are linear in the initial time value of log EP, and their variance converges to zero with rate of convergence equal to minus one. This means that, in our model, the stock prices dynamics may generate bubbles and crashes in the short and medium run, nevertheless the averaged earnings-price ratio is a good predictor of future long-run returns, as claimed by Campbell and Shiller (1988a), Shiller (2000), Lander et al. (1997). The result holds for both returns and gross returns; in the latter case we assume that the log dividend-to-price ratio follows a stationary stochastic process as in Campbell and Shiller (1988a,b). Relaxing this assumption Engsted et al. (2012) have recently investigated a model where explosive bubbles may arise, while Campbell (2008) has set forth an innovative approach to estimate the equity risk premium.

In section 2 we review the econometric issues arising in when a rate of return is regressed on

a lagged stochastic regressor where the regression disturbance is correlated with the regressor's innovation. In section 3 we describe the data set we are going to use in the estimation of the model. Section 4 sets up the model and derives the theoretical results. In section 5 we estimate the parameters of our model for both the Standard and Poor Composite Stock Price Index and the NYSE/AMEX value weighted index, and we present the results of a Monte Carlo simulation, which confirms the ability of the proposed model to capture the relevant features of the historical time series. Section 6 concludes. Technical proofs are gathered in the appendix.

2 Econometric framework

The approach that we discuss in this paper is based on the advances achieved in recent years by different authors to deal with persistent regressions. Although the methodology that we employ is of general applicability, we detail it with a specific focus on the topic of index returns predictability with respect to valuation ratios.

We refer to the real (inflation adjusted) price of the stock index, measured at the beginning of time period t , with P_t , while the notation D_t corresponds to the real dividend paid on the portfolio during the period between t and $t + 1$. Accordingly, we write the real log gross return on the index held from the beginning of time t and the beginning of time $t + 1$ as

$$H_t = \log(P_{t+1} + D_t) - \log P_t.$$

We provide a description of the return dynamics on a monthly basis. Therefore, the notation $t + 1$ refers to the time instant t increased by one month and the real gross yield over a period of length h months corresponds to

$$y_{t,h} = \frac{1}{h} \sum_{i=0}^{h-1} H_{t+i}. \tag{1}$$

We also introduce the index log price $p_t = \log P_t$, in terms of which the gross yield can be

rewritten as

$$y_{t,h} = \frac{1}{h} \sum_{i=0}^{h-1} (p_{t+i+1} - p_{t+i}) + \frac{1}{h} \sum_{i=0}^{h-1} \log \left(1 + \frac{D_{t+i}}{P_{t+1+i}} \right), \quad (2)$$

where the first term on the right hand side is equivalent to $(p_{t+h} - p_t)/h$. The dependent variable that we study throughout the paper is the gross return of the stock index, while the regressor that we consider is the log Earnings-to-Price ratio (log EP) $x_t \doteq \log \langle e \rangle_t^{10} - p_t$. The quantity $\langle e \rangle_t^{10}$ refers to a moving average of real earnings over a time window of ten years. At variance with Campbell and Shiller (1988b), where the analysis is based on geometrical averages, we employ arithmetical averages. The very reason is that in this way our study can be extended to the case of negative earnings. Indeed, while it is quite unrealistic that an average over ten years results in a negative value, however it is not so uncommon to deal with spot negative earnings. In this latter case the definition of $\log e_t$ would be problematic. Moreover, we expect that switching from one average to the other does not significantly change the overall picture. The use of an average of earnings in computing the ratios has been strongly pushed by the literature in recognition of the cyclical variability of earnings. In Graham et al. (1962) the authors recommend an approach that “Shifts the original point of departure, or basis of computation, from the current earnings to the average earnings, which should cover a period of not less than five years, and preferably seven to ten years.”

We assume that the return y_t ³ and the predictor satisfy the relations which follow

$$y_t = \alpha + \beta x_{t-1} + u_t, \quad (3)$$

$$x_t = \theta + \rho x_{t-1} + v_t, \quad (4)$$

where the errors (u_t, v_t) are serially independent and identically distributed as bivariate normal, with zero mean and covariance matrix Σ . We follow the same formal model of Stambaugh (1999) where the time series of returns $\{y_t\}$ for $t = 1, \dots, n$ is to be predicted by a scalar first-order autoregressive time series $\{x_t\}$ for $t = 0, \dots, n - 1$. When the regressions of the above type involve financial ratios the Ordinary Least-Squares (OLS) estimator for α and β is

³To ease the notation from the remaining of this section we express the log return $y_{t-1,1}$ as y_t .

biased in finite samples. This problem is pointed out by many authors, see for instance Mankiw and Shapiro (1986), Stambaugh (1986), and is analyzed by Stambaugh (1999) who develops an expression for the bias of the estimated prediction coefficient. In Stambaugh’s words “An assumption that typically fails to hold (...) is that u_t has zero expectation conditional on $\{\dots, x_{t-1}, x_t, x_{t+1}, \dots\}$, and this is the assumption used to obtain finite-sample results in the standard setting.” Stated in different words, the behaviour of u_t is mainly determined by the values p_{t-1} and p_t so $\mathbb{E}(u_t|x_t, x_{t-1}) \neq 0$. Moreover, the dynamics of x_t is strongly persistent and largely influenced by the behaviour of p_t . When the log price p_{t-1} increases, the idiosyncratic term u_t increases too, while v_t decreases; the converse is true when the price decreases. Empirically the value of ρ is relatively close to one and as an overall effect this induces a strong anti-correlation between u_t and v_t . When both the endogeneity of the regression and the persistence of x_t are taken into account the high statistical significance of the standard OLS estimator, $\hat{\beta}^{\text{OLS}}$, found in Campbell and Shiller (1987, 1988a,b) is to be questioned. However, starting from Stambaugh (1999), Lewellen (2004) finds some evidence for predictability with valuation ratios. This result relies on the upper bound for the bias in $\hat{\beta}$ which derives from the conservative assumption that the autoregressive coefficient is very close to one. We follow the alternative approach discussed in Amihud and Hurvich (2004). Replacing the predictive regression (3) with the augmented regression

$$y_t = \alpha^c + \beta^c x_{t-1} + \phi^c v_t^c + e_t, \quad (5)$$

where

$$v_t^c = x_t - \hat{\theta}^c - \hat{\rho}^c x_{t-1}, \quad (6)$$

it is possible to show that the OLS estimator of β^c is a low-biased estimator of the true parameter β appearing in (3). The quantities $\hat{\theta}^c$ and $\hat{\rho}^c$ needed to define the regressor v_t^c correspond to reduced-bias estimators of θ and ρ , respectively. A possible choice for $\hat{\rho}^c$ corresponds to the Kendall’s low-bias estimator of the AR(1) coefficient discussed in Kendall (1954). At variance we employ the second-order bias corrected estimator proposed in Amihud and Hurvich (2004)

which reads

$$\hat{\rho}^c = \hat{\rho}^{\text{OLS}} + \frac{1 + 3\hat{\rho}^{\text{OLS}}}{n} + 3\frac{1 + 3\hat{\rho}^{\text{OLS}}}{n^2}, \quad (7)$$

where $\hat{\rho}^{\text{OLS}}$ is the standard OLS estimator. Coherently, the following relation holds

$$\hat{\theta}^c = (1 - \hat{\rho}^c) \frac{\sum_{t=0}^{n-1} x_t}{n}, \quad (8)$$

with n equal to the number of available observations. The statistical significance of β can be tested by means of a low-bias finite sample approximation of the standard error of $\hat{\beta}^c$ which deviates from the error provided by the OLS regression. The reason for this deviation is that the OLS estimate fails to take into account the additional variability of $\hat{\beta}^c$ due to the variability of $\hat{\rho}^c$. Unfortunately, within the same framework it seems not possible to derive an expression for the error associated with α . Testing their approach on a data sample of monthly records from the NYSE inflation-adjusted value-weighted index ranging from May 1963 to December 1994, Amihud and Hurvich find a highly statistically significant predictability for dividend yields. In the present paper we perform the estimation of the α and β coefficients exploiting the same methodology. However, as far as testing is concerned, we do not exclusively rely on the t -statistic but we resort to a bootstrap approach which does not rely on any distributional assumption for the residual components u_t and v_t . As we will show in section 5 we find coherent results which support the forecasting ability of the logarithmic CAPE for both the Standard and Poor and NYSE/AMEX market returns.

Before we move to the description of the bootstrap algorithm, it is worth to mention some interesting alternative results dealing with the subtle issue of predictability with valuation ratios. To the best of our knowledge, the most recent advance corresponds to Hjalmarsson (2011) which is inspired by the testing framework discussed in Campbell and Yogo (2005, 2006). Both approaches are grounded on the quasi-unit root approximation for the AR(1) coefficient of equation 4, but while the latter focuses on short-run horizons the former extends the results to arbitrary horizons. The extension to the long-run predictability involves a new element that contributes to increase the complexity of the econometric issue. Indeed, long-run performances

can be evaluated only aggregating returns over smaller time scales. Given the intrinsic scarcity of data points, the sample can be enlarged only resorting to overlapping realizations and inducing sizable serial correlation among the residual terms. This side-effect can be managed by means of auto-correlation robust estimators, for instance those discussed in Hansen and Hodrick (1980), Newey and West (1987), however, when the correlation is particularly strong, these methods show poor small sample properties as discussed in Hodrick (1992), Nelson and Kim (1993). On the other hand the new asymptotic results developed in Hjalmarsson (2011) are based on the scaling properties of the t -statistics and are tested on a sub sample of the data used in Campbell and Yogo (2006) providing consistent results and evidence of the predictive power of the logarithmic CAPE.

The bootstrap in the context of returns predictability is discussed in various papers, Kilian (1999), Goyal and Santa-Clara (2003), Welch and Goyal (2008), and more specifically for regressions based on valuation ratios in Maio (2012). The testing environment that we design goes through the following steps:

- i) We estimate the bias-corrected $\hat{\rho}^c$ and $\hat{\theta}^c$ by means of (7) and (8). Using both values we compute (6) and perform the augmented regression (5). We store the values $\hat{\alpha}^c$ and $\hat{\beta}^c$ that we find via OLS.
- ii) We consider the regression (3) and (4) under the null hypothesis $\beta = 0$ and we estimate the coefficients via OLS. We save the values $\hat{\alpha}$, $\hat{\theta}$, and $\hat{\rho}$ together with the sequence of bivariate residuals (u_t, v_t) for $t = 1, \dots, n$.
- iii) We sample $b = 1, \dots, 10^4$ bivariate sequences (u_t^b, v_t^b) with $t = s_1^b, \dots, s_n^b$. The time indices s_1^b, \dots, s_n^b are sampled with replacement from the original set of indices $t = 1, \dots, n$.
- iv) For each replication $b = 1, \dots, 10^4$, we compute a bootstrap sample by imposing the null hypothesis

$$y_t^b = \hat{\alpha} + u_t^b,$$

$$x_t^b = \hat{\theta} + \hat{\rho}x_{t-1}^b + v_t^b.$$

The initial point x_0^b is drawn uniformly from the set $\{x_t\}_{t=0,\dots,n-1}$.

- v) For each bootstrap sample we estimate the bias-corrected $\hat{\theta}^{c,b}$ and $\hat{\rho}^{c,b}$, we compute the proxy for the error terms $v_t^{c,b}$ as

$$v_t^{c,b} = x_t^b - \hat{\theta}^{c,b} - \hat{\rho}^{c,b} x_{t-1}^b,$$

and we estimate the following augmented regression via OLS

$$y_t^b = \alpha^{c,b} + \beta^{c,b} x_{t-1}^b + \phi^{c,b} v_t^{c,b} + e_t^b.$$

Finally we obtain a sample $\{\hat{\beta}^{c,b}\}$ for $b = 1, \dots, 10^4$.

- vi) The one sided p -value associated with the prediction coefficient is calculated as

$$p - \text{value} = \frac{\#\{\hat{\beta}^{c,b} \geq \hat{\beta}^c\}}{10^4},$$

where the numerator on the right hand side denotes the number of bootstrap estimates higher than the original value, and we implicitly assume that $\hat{\beta}^c > 0$.

3 Data set

The data set analyzed in this paper consists of records on a monthly basis of prices, earnings, and dividends for two stock price indexes for the US market. The first is the Standard and Poor Composite Stock Price Index (S&P) data discussed in Campbell and Shiller (1987, 1988a,b), and freely available from Robert J. Shiller's webpage <http://www.econ.yale.edu/>. These time series cover the entire period from January 1871 until December 2012. The second is NYSE/AMEX value weighted index data (1926-1994) from the Center for Research in Security Prices (CRSP) available from J. Campbell home page <http://scholar.harvard.edu/campbell/data>. Since earnings data are not available for CRSP series we use the corresponding S&P earnings-to-price ratios from Shiller. Figure 1 in Campbell and Yogo (2006) compares the time series of the log dividend-to-price ratio for the NYSE/AMEX value-weighted index and the log smoothed earnings-price for the S&P and provides support for our choice.

4 The model

In the current section we characterise a simple dynamical system driven both by a value investment and a momentum component. In the first instance we consider only the role played by stock prices, but later we take into account gross returns and the effect of dividends. Our model rests on the two assumptions which follow.

Assumption 1 *Financial returns are naturally driven by the growth rate μ_t , and by an ancillary process ξ_t satisfying*

$$\xi_{t+1} = \xi_t + \frac{\kappa}{1-\gamma} \sigma_p W_t^p, \quad (9)$$

with initial time condition $\xi_0 = 0$. The ξ_t process ensures the diffusive behaviour of stock prices. As will be clear later, the positive parameter σ_p corresponds to the long run returns volatility, while the prefactor depending on $\kappa > 0$ and $0 < \gamma < 1$ has been added for future convenience. The W_t^p 's for $t = 0, \dots, h$ are independent identically distributed (i.i.d.) standard Gaussian increments.

Assumption 2 *The dynamics of log returns' growth rate is given in terms of the difference equation*

$$\mu_{t+1} = \gamma \mu_t + \kappa \left[\log \langle e \rangle_t^{10} - p_t + \mathcal{H} + g \mathcal{F} \left(\log \langle e \rangle_0^{10} - p_0 \right) t \right] + \sigma_\mu W_t^\mu \quad (10)$$

where the parameter σ_μ is a positive volatility constant, and the W_t^μ 's for $t = 0, \dots, h$ are i.i.d. standard Gaussian increments, not dependent on the W_t^p 's. The return growth μ_t depends on its own value in the previous period, with γ agents' sensitivity to market trend. This momentum effect can be naturally justified in terms of agents' expectation that returns are higher in bullish markets than in bearish ones. The "fundamental" component is proportional to a function \mathcal{F} of the level of the logarithmic averaged earnings-to-price ratio at time zero, where the proportionality constant is given by the rate of earnings growth g , implicitly defined by $\langle e \rangle_t^{10} = \langle e \rangle_0^{10} \exp(gt)$. This component grows linearly with time, but we also allow for a correction equal to \mathcal{H} , possibly function of the initial time log EP ratio. The initial time value of log EP determines the refer-

ence growth level, from which the actual stock price may deviate as an effect of random external disturbances. Investors reallocate assets in response to this disequilibrium causing stock prices to move in the direction that reduces the deviation. The full adjustment is not immediate but it takes a typical time κ^{-1} the price to mean revert to the fundamental level.

In conclusion, our model states the dynamics of log prices can be fully specified by the linear stochastic difference system

$$\begin{cases} p_{t+1} &= p_t + \mu_t + \xi_t \\ \mu_{t+1} &= \gamma\mu_t + \kappa(\log\langle e \rangle_t^{10} - p_t + \mathcal{H} + g\mathcal{F}t) + \sigma_\mu W_t^\mu, \\ \xi_{t+1} &= \xi_t + \frac{\kappa}{1-\gamma}\sigma_p W_t^p \end{cases}, \quad (11)$$

with $p_0 = \log P_0$ and μ_0 initial time conditions. By a further change of variable $Y_t = p_t - \log\langle e \rangle_0^{10}$, the expected values $\mathbb{E}_0[Y_t]$, $\mathbb{E}_0[\mu_t]$, $\mathbb{E}_0[\xi_t]$ conditional at the information available at time 0 evolve according to the linear first order difference system

$$\begin{cases} \mathbb{E}_0[Y_{t+1}] &= \mathbb{E}_0[Y_t] + \mathbb{E}_0[\mu_t] + \mathbb{E}_0[\xi_t] \\ \mathbb{E}_0[\mu_{t+1}] &= -\kappa\mathbb{E}_0[Y_t] + \gamma\mathbb{E}_0[\mu_t] + \kappa\mathcal{H} + \kappa g(1 + \mathcal{F})t, \\ \mathbb{E}_0[\xi_{t+1}] &= \mathbb{E}_0[\xi_t] \end{cases},$$

with the initial conditions

$$\begin{cases} \mathbb{E}_0[Y_0] &= p_0 - \log\langle e \rangle_0^{10} \\ \mathbb{E}_0[\mu_0] &= \mu_0 \\ \mathbb{E}_0[\xi_0] &= 0 \end{cases}.$$

Lemma 1. *The expected rate of return is linear in $\mathcal{F}(\log\langle e \rangle_0^{10} - p_0)$, provided we consider a sufficiently large number of periods h*

$$\mathbb{E}_0 \left[\frac{1}{h} \log \frac{P_{0+h}}{P_0} \right] = g(1 + \mathcal{F}) + O\left(\frac{1}{h}\right). \quad (12)$$

Moreover its variance converges to zero with rate of convergence equal to minus one

$$\text{Var}_0 \left[\frac{1}{h} \log \frac{P_{0+h}}{P_0} \right] = \frac{\sigma_p^2}{h} + o\left(\frac{1}{h}\right). \quad (13)$$

A posteriori it is clear why we fixed the proportionality constant of the fundamental component equal to g . Indeed, in case g were zero not only averaged earnings but also stock returns would grow sub-exponentially. The variance of Y_h increases linearly with h , the diffusive behaviour of stock returns that any realistic model should satisfy in a first instance. Finer effects, like possible return non zero autocorrelations induced by business cycles over long horizons could be included in our model modifying the covariance structure of the $\{W_t^p\}$ noise. In this respect it is worth to remark that none of our results crucially rest on the Gaussian assumption for the noise increments.

From previous Lemma the result that we state below follows.

Corollary 2. *If \mathcal{F} is linear in $\log\langle e \rangle_0^{10} - p_0$, than the model (11) is able to reproduce the linear scaling of returns with the initial log EP ratio on the long run.*

The next step forward is to add a dividend components to our stock returns model. In this respect we mainly follow the proposal discussed in Campbell and Shiller (1988a,b), where it is argued that the log dividend-to-price ratio follows a stationary stochastic process, and the fixed mean of $\log D_t - \log P_t$, $\log \mathcal{G}$, can be used as an expansion point

$$\Delta(d_{t-1} - p_t) = -\theta(d_{t-1} - p_t - \log \mathcal{G}) + \sigma_d W_t^d, \quad (14)$$

with $d_t = \log D_t$, $\sigma_d > 0$, and $\{W_t^d\}$ for $t = 1, \dots, h$ i.i.d. standard Gaussian variates. At variance with the proposal of Campbell and Shiller, we also assume that \mathcal{G} can possibly depend on the log EP at time zero, i.e. $\mathcal{G} = \mathcal{G}(\log\langle e \rangle_0^{10} - p_0)$.

Lemma 3. *In the limit $h \gg 1$ the quantity $\frac{1}{h} \sum_{i=0}^{h-1} \log \left(1 + \frac{D_i}{P_{i+1}} \right)$ contributes to the rate of log gross returns proportionally to \mathcal{G} , and its variance converges to zero with rate minus one.*

Corollary 4. *If \mathcal{G} is linear in $\log\langle e \rangle_0^{10} - p_0$, than equation (14) is able to reproduce the linear scaling of dividends' contribution to the yield with the initial log EP ratio on the long run.*

We are now ready to state the main theoretical result of this paper.

Proposition 5. *The expected gross yield (1) is linear in \mathcal{F} and \mathcal{G} , provided we consider sufficiently long horizons*

$$\mathbb{E}_0[y_h] \simeq g(1 + \mathcal{F}) + \mathcal{G} + O\left(\frac{1}{h}\right). \quad (15)$$

The proof is an immediate consequence of the Corollaries 2 and 4. Interestingly from the proofs of previous Lemmas given in Appendix (A) we can explicitly identify the contribution to the long term yield whose scaling over time is more persistent, and which ultimately determines its convergence towards the limit $g(1 + \mathcal{F}) + \mathcal{G}$. The expression of the leading correction proportional to one over h reads

$$\mathcal{H} - g(1 + \mathcal{F}) \left[1 + \kappa \frac{1 - \lambda_- \lambda_+}{(1 - \lambda_-)^2 (1 - \lambda_+)^2} \right] + \log EP_0 + \mathcal{G} \left(1 - \frac{1}{\theta} \right) (\log DP_0 - \log \mathcal{G}). \quad (16)$$

From equation (21) we know that $(1 - \theta)^t$ plays the role of a damping function. Assuming $\theta \ll 1$, it reduces to $\exp(-t/\tau_\theta)$ where the quantity $\tau_\theta = 1/\theta$ plays the role of the typical time scale of the mean-reverting process. We reasonably expect that the last term in the expression (16) does not contribute too much to the leading correction, since we neither expect an extremely large time scale for the process nor an extreme discrepancy between $\log DP_0$ and $\log \mathcal{G}$. Another interesting point to note is that the choice of μ_0 plays a minor role in relation to the speed of convergence of the process to the long term expected value. Indeed, μ_0 does not appear in expression (16), while from (19) we know that its effect is exponentially damped by λ_-^h and λ_+^h . Being $0 < \lambda_- < \lambda_+ < 1$, the dominating contribution for $h \gg 1$ is given by λ_+^h , implying a typical damping scale equal to $-(\log \lambda_+)^{-1}$.

5 Parameters estimation

Corollaries 2 and 4 state that the linear scaling of long term yield with the level of log EP ratio can be reproduced by our dynamical model if we assume for the quantities $g(1 + \mathcal{F})$ and \mathcal{G} the form $g(1 + \mathcal{F}) = \alpha_{\mathcal{F}} + \beta_{\mathcal{F}} \log EP_0$ and $\mathcal{G} = \alpha_{\mathcal{G}} + \beta_{\mathcal{G}} \log EP_0$. Following the same approach of Amihud and Hurvich (2004), we estimate the values of $\alpha_{\mathcal{F}}$, $\beta_{\mathcal{F}}$ from the regression of log returns over one year horizon on the initial log EP. The quantities labelled with \mathcal{g} are obtained in the same way after regressing the dividends' contribution to the yield on the same regressors. The results that we find are reported in Table 1 for both the S&P and NYSE/AMEX indices. We also report the prediction coefficients of the gross returns on the initial logarithmic CAPE, α and β . The statistical significance of the β coefficients is assessed by means of the reduced-biased standard errors discussed in Amihud and Hurvich (2004) and according to the bootstrap

	α	β	$\alpha_{\mathcal{F}}$	$\beta_{\mathcal{F}}$	$\alpha_{\mathcal{G}}$	$\beta_{\mathcal{G}}$
S&P	3667	1023 ± 445 (0.048)	2531	767 ± 459 (0.091)	1527	393 ± 19 ($< 10^{-3}$)
NYSE/AMEX	6684	2228 ± 882 (0.024)	5681	1880 ± 840 (0.043)	1289	305 ± 31 ($< 10^{-3}$)

Table 1: All coefficient values and associated errors are expressed on a yearly basis and in 10^{-4} units. Between parentheses we report the p -values obtained by means of 10^4 bootstrap samples.

approach devised in section 2. If we compute the t -statistics for the β parameters we find $t = 2.29$ and $t = 2.52$ for the S&P and CRSP time series respectively, which imply a statistical significant evidence of predictability. The predictive coefficient values can be disaggregated in two components, one due to the net return growth and the other associated to the dividend component. The data in Table 1 show that the long-run return predictability is largely due to the highly predictable behaviour of the log dividend-to-price ratio. Nonetheless, especially for the case of the CRSP time series, also the net return component determines the predictive regression in a statistically validated way. The evidence that we observe confirms the empirical results discussed in Table 4 of Campbell and Yogo (2006). In that case it is however important to mention that the authors consider the excess gross return as dependent variable. We verified in our data that the growth in excess with respect to the long run risk free yield tends in general to better perform when we need to compute the significance of the predictive regression. In Table 1 we moreover show that the null hypothesis $\beta = 0$ is also rejected in light of our bootstrap approach which does not rely on any ad hoc normal assumption for the idiosyncratic component e_t . We sample 10^4 synthetic bootstrap copies and we measure a bootstrap p -value lower than 5% for both the S&P and NYSE/AMEX time series.

Having estimated the linear functions \mathcal{F} and \mathcal{G} by means of the above procedure, the next relevant parameter to estimate is the rate of exponential growth of averaged earnings g . This can be easily obtained iteratively, first regressing a vector of 192 monthly logarithmic averages on a linear function of h , then rolling the time window and repeating the estimate from January 1881 for the S&P index and from December 1926 for the NYSE/AMEX index until the last

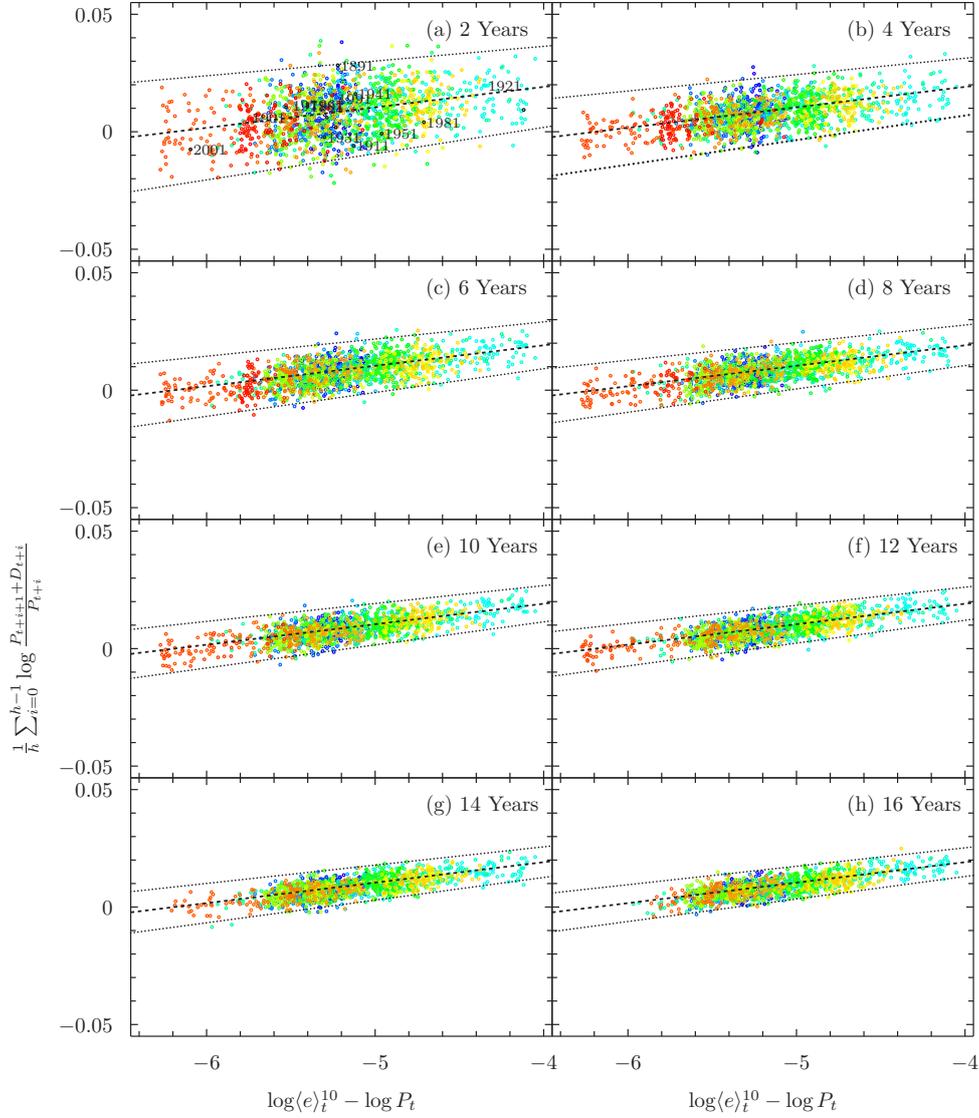


Figure 1: Monte Carlo scenarios generated simulating the model given by equations (11) and (14) with initial time conditions Y_0 , μ_0 , and $\log DP_0$ equal to the empirical ones. The dashed line corresponds to the long-run behaviour predicted by the equation (15) using the coefficient values of Table 1 for the S&P index rescaled to a monthly level. The dotted lines correspond to the boundaries of the 95% confidence level region. Points are organised in chronological order according to the color scale ranging from dark blue to red passing through light blue, green, yellow, and orange; labels in the top left panel refer to points which correspond to the first month of the specified year.

admissible date. In this way we obtain a sample of g values from which empirical 68% confidence intervals can be computed, for the results on a monthly basis from Shiller's and Campbell's data series please refer to Table 2. The value of θ and σ_d are estimated regressing $\Delta(d_{t-1} - p_t)$ on $d_{t-1} - p_t - \log \mathcal{G}$, employing the same technique of rolling windows. The measured mean value for θ is equal to 0.0271 and 0.0445 for the S&P and NYSE/AMEX index respectively, and following the consideration written at the end of the previous section these values imply typical scales for mean-reversion of order 37 and 22 months, respectively. The parameters of the μ_t process are also estimated by means of linear regression of μ_{t+1} on μ_t , and on $\log \langle e \rangle_t^{10} - p_t + \mathcal{H} + g\mathcal{F}t$. In the first instance this requires to introduce a suitable proxy for the unobservable variable μ_t . Our approach is to estimate it by means of $p_t - p_{t-1}$, but different choices are admissible, for example in terms of an exponential weighted average of past price returns. However in this latter case the undesirable dependence of the parameters on the value of the exponential weight would be introduced. Moreover \mathcal{H} is still an unknown quantity that we have to choose before performing the optimization procedure. Recalling expression (16), it is reasonable to assume it linear in the log EP ratio $\mathcal{H} = \alpha_{\mathcal{H}} + \beta_{\mathcal{H}} \log EP_0$ and to initialise the procedure with arbitrary values for the linear coefficients. The numerical process then produces optimal estimates of γ and κ , for suitable $\alpha_{\mathcal{H}}$ and $\beta_{\mathcal{H}}$ minimizing the leading correction term (16). In particular, for $\alpha_{\mathcal{H}} = 0.85$ and $\beta_{\mathcal{H}} = -0.85$ we find $\gamma = 0.25$ and $\kappa = 0.03$ for S&P, and for $\alpha_{\mathcal{H}} = 2.62$ and $\beta_{\mathcal{H}} = -0.52$ we find $\gamma = 0.08$ and $\kappa = 0.03$ for NYSE/AMEX, both satisfying the constraint $4\kappa \leq (1 - \gamma)^2$. For the estimate of σ_p we approach the problem from a different perspective. Indeed from equation (13) we know that the variance of $(p_h - p_0)/h$ scales as one over h with proportionality constant equal to σ_p^2 . We therefore fit the empirical curves with a linear relation and we find the values of 0.182% and 0.168% reported in Table 2. In Figure 1 we present the results of a Monte Carlo simulation of the model (11)-(14). Each point corresponds to a single realization of $y_{t,h}$ with $h = 24, \dots, 192$ months with t starting from January 1881. The initial time values of Y_0 , μ_0 , and $\log DP_0$ are fixed equal to the empirical ones. We plot the linear relation between yields and the log EP as predicted by the equation (15) using the coefficient values of Table 1 for the S&P index rescaled to a monthly level. We also provide the boundaries of the 95% confidence level region which we compute by means of the formulas (13) and (22). All the analytical predictions are fully confirmed by the Monte Carlo numerical results.

	S&P	NYSE/AMEX
g ($\times 10^{-4}$ <i>month</i> ⁻¹)	12 (-3, 31)	19 (4, 32)
θ ($\times 10^{-4}$ <i>month</i> ⁻¹)	271 (111, 430)	445 (197, 780)
γ	0.25 (0.18, 0.33)	0.08 (0.02, 0.15)
κ ($\times 10^{-4}$ <i>month</i> ⁻¹)	323 (81, 597)	304 (70, 547)
σ_d^2 ($\times 10^{-4}$ <i>month</i>)	13 (10, 20)	19 (13, 25)
σ_μ^2 ($\times 10^{-4}$ <i>month</i>)	12 (9, 18)	17 (12, 24)
σ_p^2 ($\times 10^{-4}$ <i>month</i>)	18.2 (18.1, 18.3)	16.8 (16.5, 17.1)

Table 2: Parameter values for the S&P and NYSE/AMEX time series with associated 68% confidence level intervals.

Finally, we test the goodness of the estimated parameters and the effectiveness of our dynamical system. In Figures (2) and (3) we plot the prediction of equation (15) with the associated confidence band on the net returns and on the gross yields, respectively, for time horizons ranging from 24 to 192 months. As expected the model captures the shrinking of the historical data cloud with a scaling exponent dominated by the diffusive component of the price dynamics. As far as the central value is concerned, the consistency is very good for the short time horizons, whilst it worsens for longer runs. This effect is partially expected, since the linear coefficients of the predictive regressions are constants quantities which are exogenous to the dynamical model (11) and do not evolve with time. We are currently working on an econometric approach to the model (3)-(4) in order to provide a sound theoretical basis to the extension of the low-bias procedure of Amihud and Hurvich (2004) to long-run horizons. Preliminary results suggest that the β coefficient rescales geometrically over time, where the basis of the geometric sequence corresponds to the AR(1) coefficient of equation (4). We plan to correct for this minor effect in a future extension of our model.

6 Conclusion and Perspectives

This paper documents predictability of stock index returns for the U.S. market on the basis of a simple valuation metric, the cyclically adjusted price-earnings ratio, and proposes a discrete time dynamical model able to capture the long-run behaviour.

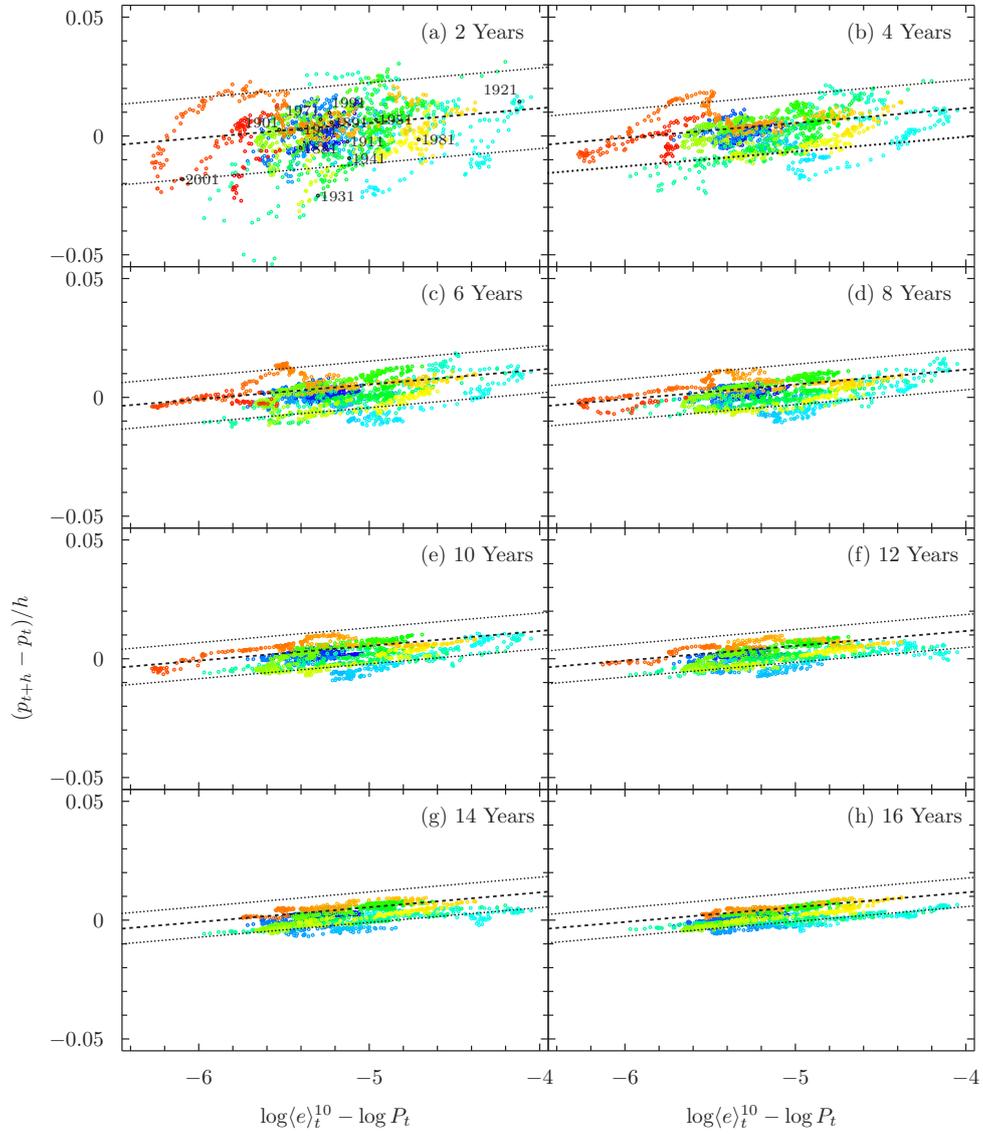


Figure 2: Regression of the yield $(p_{t+h} - p_t)/h$ on the explanatory log earning - price ratio. Solid points: empirical data, dashed line: model prediction, dotted line: 95% confidence level from model prediction. Color scale as in Figure 1.

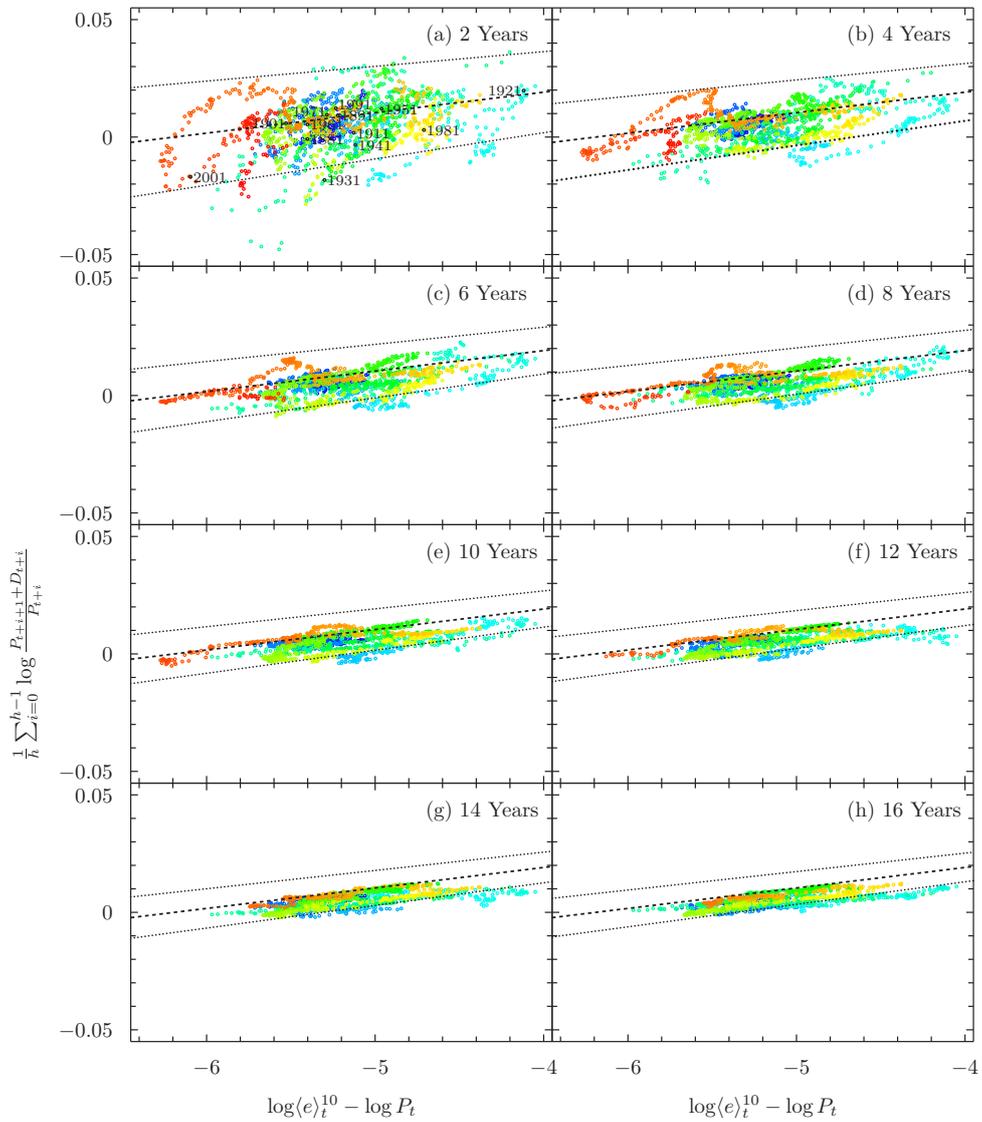


Figure 3: Regression of yield (1) on the explanatory log earning - price ratio. Points, lines, labels, and color scale as in Figure 2.

Substantial evidence of the importance of fundamentals in the valuation of international stock markets has been accumulated by the proponents of fundamental indexation Arnott et al. (2005). Practitioners and academicians alike have been using several valuation measures for estimating the intrinsic value of a stock index: for example in table 2 of Poterba and Samwick (1995) the ratio of market value of corporate stock to GDP, the year-end price-to-earnings ratio, the year-end price-to-dividend ratio and Tobin's q are reported from 1947 to 1995 in an effort of alerting the reader on the possible overvaluation of the index ⁴. In particular Tobin's q has been proposed as another efficient method of measuring the value of the stock market, with an efficiency comparable to the CAPE Smithers (2009). The q ratio is the ratio of price to net worth at replacement cost rather than the historic or book cost of companies. It therefore allows for the impact of inflation, much alike the CAPE which averages real earnings over a ten year span. It would be interesting to carry out an empirical analysis of the relationship between Tobin's q and future stock index returns as far as to extend the present approach to countries other than the U.S.. Both perspectives are worth to be followed but require high quality long-term time series which at the moment we have not been able to find. Longer time-series would also be important so as to investigate possible "evolutionary" phenomena like the end of the relationship between market valuation and interest rates, which may perhaps be interpreted as an example of the adaptive market hypothesis Farmer and Lo (1999), Lo (2004).

References

- Yakov Amihud and Clifford M. Hurvich. Predictive regressions: A reduced-bias estimation method. *Journal of Financial and Quantitative Analysis*, 39(04):813–841, 2004.
- Robert D. Arnott, Jason Hsu, and Philip Moore. Fundamental indexation. *Financial Analysts Journal*, 61(2):83–99, 2005.

⁴"It is difficult to distill a simple conclusion from table 2. While price-to-earnings ratios are not unusually high at present, other measures of stock price valuation are at, or near, historical highs. (...) Table 2 does suggest, however, that in assessing the macroeconomic consequences of stock price movements, it may be important to distinguish between stock price fluctuations that are associated with movements in the price-to-earnings or price-to-dividends ratios and those that are not. A number of recent studies suggest that variations in the earnings-to-price ratio are correlated with prospective stock market returns, (...). Sharp increases in either the price-to-earnings or the price-to-dividends ratio, other things equal, are associated with lower prospective returns."

- John C. Campbell. Viewpoint: Estimating the equity premium. *Canadian Journal of Economics / Revue Canadienne d'Economique*, 41(1):1–21, 2008.
- John Y. Campbell and Robert J. Shiller. Cointegration and tests of present value models. *Journal of Political Economy*, 95(5):1062–1088, 1987.
- John Y. Campbell and Robert J. Shiller. The dividend-price ratio and expectations of future dividends and discount factors. *Review of Financial Studies*, 1(3):195–228, 1988a.
- John Y. Campbell and Robert J. Shiller. Stock prices, earnings and expected dividends. *Journal of Finance*, XLIII(3):661–676, 1988b.
- John Y. Campbell and Robert J. Shiller. Valuation ratios and the long-run stock market outlook. *The Journal of Portfolio Management*, 24(2):11–26, 1998.
- John Y. Campbell and Motohiro Yogo. Implementing the econometric methods in efficient tests of stock return predictability. *Unpublished working paper. University of Pennsylvania*, 2005.
- John Y. Campbell and Motohiro Yogo. Efficient tests of stock return predictability. *Journal of financial economics*, 81(1):27–60, 2006.
- Carl Chiarella. The dynamics of speculative behaviour. *Annals of Operations Research*, 37(1):101–123, 1992.
- Tom Engsted, Thomas Q. Pedersen, and Carsten Tanggaard. The log-linear return approximation, bubbles, and predictability. *Journal of Financial and Quantitative Analysis*, 47:643–665, 2012.
- Eugene F. Fama. The behavior of stock-market prices. *The Journal of Business*, 38:34–105, 1965.
- Eugene F. Fama. Efficient capital markets: A review of theory and empirical work. *Journal of Finance*, 25(2):383–417, 1970.
- Doyne Farmer and Andrew W. Lo. Frontiers of finance: Evolution and efficient markets. *Proceedings of the National Academy of Sciences*, 96(18):9991–9992, 1999.

- Amit Goyal and Pedro Santa-Clara. Idiosyncratic risk matters! *The Journal of Finance*, 58 (3):975–1008, 2003.
- Benjamin Graham and Spencer B. Meredith. *The Interpretation of Financial Statements: The Classic 1937 Edition*. HarperBusiness, 1998.
- Benjamin Graham, David L. Dodd, and Sidney Cottle. *Security Analysis*. McGraw-Hill, New York, 1962.
- Lars P. Hansen and Robert J. Hodrick. Forward exchange rates as optimal predictors of future spot rates: An econometric analysis. *The Journal of Political Economy*, pages 829–853, 1980.
- Robert J. Hodrick. Dividend yields and expected stock returns: Alternative procedures for inference and measurement. *Review of Financial studies*, 5(3):357–386, 1992.
- Erik Hjalmarsson. New methods for inference in long-horizon regressions. *Journal of Financial and Quantitative Analysis*, 46(3):815, 2011.
- Maurice G. Kendall. Note on bias in the estimation of autocorrelation. *Biometrika*, 41(3-4): 403–404, 1954.
- Lutz Kilian. Exchange rates and monetary fundamentals: what do we learn from long-horizon regressions? *Journal of Applied Econometrics*, 14(5):491–510, 1999.
- Joel Lander, Athanasios Orphanides, and Martha Douvogiannis. Earnings forecasts and the predictability of stock returns: Evidence from trading the S&P. Technical report, Board of Governors of the Federal Reserve System, 1997.
- Jonathan Lewellen. Predicting returns with financial ratios. *Journal of Financial Economics*, 74(2):209–235, 2004.
- Andrew W. Lo. The adaptive markets hypothesis: Market efficiency from an evolutionary perspective. *Journal of Portfolio Management*, 30(5):15–29, 2004.
- Paulo Maio. The fed model and the predictability of stock returns. *Review of Finance*, 2012. doi: 10.1093/rof/rfs025.

- Benoit Mandelbrot. The variation of certain speculative prices. *The Journal of Business*, 36 (4):394–419, 1965.
- Gregory N. Mankiw and Matthew D. Shapiro. Do we reject too often? : Small sample properties of tests of rational expectations models. *Economics Letters*, 20(2):139–145, 1986.
- James Myers and Bhaskaran Swaminathan. What is the intrinsic Value of the Dow? *Journal of Finance*, 54(5):1693–1741, 1999.
- Charles R. Nelson and Myung J. Kim. Predictable stock returns: The role of small sample bias. *The Journal of Finance*, 48(2):641–661, 1993.
- Whitney K. Newey and Kenneth D. West. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica: Journal of the Econometric Society*, pages 703–708, 1987.
- James M. Poterba and Andrew A. Samwick. Stock ownership patterns, stock market fluctuations, and consumption. *Brookings Papers on Economic Activity*, 1995(2):295–372, 1995.
- Paul Samuelson. Proof that properly anticipated prices fluctuate randomly. *Industrial Management Review*, 6(2):41–49, 1965.
- Robert J. Shiller. *Irrational Exuberance*. Princeton University Press, 2000.
- Andrew Smithers. *Wall Street Revalued. Imperfect Markets and Inept Central Bankers*. John Wiley and Sons, Ltd, 2009.
- Robert F. Stambaugh. *Bias in regressions with lagged stochastic regressors*. Center for Research in Security Prices, Graduate School of Business, University of Chicago, 1986.
- Robert F. Stambaugh. Predictive regressions. *Journal of Financial Economics*, 54(3):375–421, 1999.
- Robert A. Weigand and Robert Irons. Does the market P/E ratio revert back to “average”? *Investment Management and Financial Innovations*, 3(3):30–39, 2006.
- Ivo Welch and Amit Goyal. A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, 21(4):1455–1508, 2008.

A Proofs

Proof of Lemma 1. The stochastic dynamical system (11) may be put in the vector form

$$\mathbf{V}_{t+1} = \mathbb{J}\mathbf{V}_t + \kappa\mathcal{H} \mathbf{e}_2 + \kappa g(1 + \mathcal{F})t \mathbf{e}_2 + \sigma_\mu W_t^\mu \mathbf{e}_2 + \frac{\kappa}{1-\gamma} \sigma_p W_t^p \mathbf{e}_3 \quad (17)$$

where $\mathbb{J} = \begin{bmatrix} 1 & 1 & 1 \\ -\kappa & \gamma & 0 \\ 0 & 0 & 1 \end{bmatrix}$ is the matrix of the systems, \mathbf{e}_i for $i = 1, 2, 3$ is the orthonormal canonical base, and $\mathbf{V}_t = \begin{bmatrix} Y_t \\ \mu_t \\ \xi_t \end{bmatrix}$. The set of the eigenvalues of \mathbb{J} includes 1, independently of κ and γ , and two ones more

$$\begin{aligned} \lambda_+ &= \frac{\gamma+1}{2} + \frac{1}{2}\sqrt{(1-\gamma)^2 - 4\kappa}, \\ \lambda_- &= \frac{\gamma+1}{2} - \frac{1}{2}\sqrt{(1-\gamma)^2 - 4\kappa}. \end{aligned}$$

In order to ensure $|\lambda_+| < 1$ and $|\lambda_-| < 1$, we need to impose the following restrictions

$$0 < \gamma < 1, \quad \text{and} \quad 0 < \kappa \leq \frac{(1-\gamma)^2}{4}. \quad (18)$$

Iterating (17) we obtain

$$\begin{aligned} \mathbf{V}_h &= \mathbb{J}^h \mathbf{V}_0 + \kappa [g(1 + \mathcal{F})(h-1) + \mathcal{H}] \sum_{k=0}^{h-1} \mathbb{J}^k \mathbf{e}_2 - \kappa g(1 + \mathcal{F}) \sum_{k=0}^{h-1} k \mathbb{J}^k \mathbf{e}_2 \\ &\quad + \sigma_\mu \sum_{k=0}^{h-1} W_k^\mu \mathbb{J}^{h-1-k} \mathbf{e}_2 + \frac{\kappa}{1-\gamma} \sigma_p \sum_{k=0}^{h-1} W_k^p \mathbb{J}^{h-1-k} \mathbf{e}_3. \end{aligned}$$

We can rewrite more conveniently $\mathbb{J} = \mathbb{Q}\Lambda\mathbb{Q}^{-1}$ by means of the matrices $\Lambda = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \lambda_+ & 0 \\ 0 & 0 & \lambda_- \end{bmatrix}$,

$$\mathbb{Q} = \begin{bmatrix} 1 - \gamma & 1 & 1 \\ -\kappa & \lambda_+ - 1 & \lambda_- - 1 \\ \kappa & 0 & 0 \end{bmatrix}, \text{ and } \mathbb{Q}^{-1} = \frac{1}{|\mathbb{Q}|} \begin{bmatrix} 0 & 0 & \lambda_- - \lambda_+ \\ \kappa(\lambda_- - 1) & -\kappa & (1 - \gamma)(1 - \lambda_-) - \kappa \\ \kappa(1 - \lambda_+) & \kappa & \kappa - (1 - \gamma)(1 - \lambda_+) \end{bmatrix}$$

with $|\mathbb{Q}| = \kappa(\lambda_- - \lambda_+)$. Taking expectation of $\mathbf{e}_1^t \mathbf{V}_h$ we obtain

$$\begin{aligned} \mathbb{E}_0 [Y_h] &= \mathbf{e}_1^t \mathbb{Q} \Lambda^h \mathbb{Q}^{-1} \mathbf{V}_0 \\ &+ \kappa [g(1 + \mathcal{F})(h - 1) + \mathcal{H}] \sum_{k=0}^{h-1} \mathbf{e}_1^t \mathbb{Q} \Lambda^k \mathbb{Q}^{-1} \mathbf{e}_2 \\ &- \kappa g(1 + \mathcal{F}) \sum_{k=0}^{h-1} k \mathbf{e}_1^t \mathbb{Q} \Lambda^k \mathbb{Q}^{-1} \mathbf{e}_2 \\ &= -\frac{\lambda_+^h}{\lambda_- - \lambda_+} [(1 - \lambda_-)Y_0 + \mu_0] + \frac{\lambda_-^h}{\lambda_- - \lambda_+} [(1 - \lambda_+)Y_0 + \mu_0] \\ &+ g(1 + \mathcal{F})(h - 1) + \mathcal{H} \\ &- [g(1 + \mathcal{F}) - \mathcal{H}] \frac{\kappa}{\lambda_- - \lambda_+} \left(\frac{\lambda_+^h}{1 - \lambda_+} - \frac{\lambda_-^h}{1 - \lambda_-} \right) \\ &- g(1 + \mathcal{F}) \frac{\kappa}{\lambda_- - \lambda_+} \left[\lambda_- \frac{1 - \lambda_-^h}{(1 - \lambda_-)^2} - \lambda_+ \frac{1 - \lambda_+^h}{(1 - \lambda_+)^2} \right]. \end{aligned} \quad (19)$$

From previous expression relation (12) immediately follows.

As far as the variance of Y_h is concerned, we have

$$Var_0 [Y_h] = \sum_{k=0}^{h-1} \left[\sigma_\mu^2 (\mathbf{e}_1^t \mathbb{Q} \Lambda^k \mathbb{Q}^{-1} \mathbf{e}_2)^2 + \frac{\kappa^2}{(1 - \gamma)^2} \sigma_p^2 (\mathbf{e}_1^t \mathbb{Q} \Lambda^k \mathbb{Q}^{-1} \mathbf{e}_3)^2 \right], \quad (20)$$

from which we find

$$\begin{aligned} Var_0 [Y_h] &= \frac{\sigma_\mu^2}{(\lambda_- - \lambda_+)^2} \sum_{k=0}^{h-1} [\lambda_-^k - \lambda_+^k]^2 \\ &+ \frac{\sigma_p^2}{(\lambda_- - \lambda_+)^2} \sum_{k=0}^{h-1} \left[\lambda_- - \lambda_+ + \lambda_+^k (1 - \lambda_- - \frac{\kappa}{1 - \gamma}) \right. \\ &\left. + \lambda_-^k (\frac{\kappa}{1 - \gamma} - 1 + \lambda_+) \right]^2, \end{aligned}$$

and the thesis (13) follows. \square

Proof of Lemma 3. On a monthly basis the quantity D_t/P_{t+1} is of order 10^{-3} , and we are

allowed to replace $\log\left(1 + \frac{D_t}{P_{t+1}}\right)$ with D_t/P_{t+1} . In the same spirit of Campbell and Shiller (1988a,b), in order to prove the thesis we need to Taylor expand the quantities of interest around $\log \mathcal{G}$. In particular, we have

$$\frac{D_t}{P_{t+1}} = \exp(d_t - p_{t+1}) \simeq \exp(\log \mathcal{G}) (1 + d_t - p_{t+1} - \log \mathcal{G}) .$$

Iterating equation (14) and taking expectation, we obtain for $t \geq 0$

$$\mathbb{E}_0[d_{t-1} - p_t] = (1 - \theta)^t (d_{-1} - p_0) + \log \mathcal{G} [1 - (1 - \theta)^t] , \quad (21)$$

where d_{-1} is the log dividend prevailing at time zero being cumulated between time -1 and zero. Under the constraint $0 < \theta < 2$, we can conclude

$$\begin{aligned} \frac{1}{h} \mathbb{E}_0 \left[\sum_{i=0}^{h-1} \log \left(1 + \frac{D_i}{P_{i+1}} \right) \right] &\simeq \mathcal{G}(1 - \log \mathcal{G}) + \mathcal{G} \log \mathcal{G} \\ &\quad + \mathcal{G} \frac{1 - \theta}{\theta} (\log DP_0 - \log \mathcal{G}) \frac{1 - (1 - \theta)^h}{h} \\ &= \mathcal{G} + O\left(\frac{1}{h}\right) , \end{aligned}$$

and

$$\text{Var}_0 \left[\frac{1}{h} \sum_{i=0}^{h-1} \log \left(1 + \frac{D_i}{P_{i+1}} \right) \right] \simeq \frac{1}{h} \frac{\mathcal{G}^2 \sigma_d^2}{\theta(2 - \theta)} + o\left(\frac{1}{h}\right) . \quad (22)$$

□

Remark. The expansion performed in the above proof is similar to that proposed in Campbell and Shiller (1988a,b), where $\log(P_t + D_t)$ is approximated by $\vartheta p_t + (1 - \vartheta) \log D_t + \varkappa$ with $\vartheta = 1/(1 + \mathcal{G})$ and $\varkappa = \log(1 + \mathcal{G}) - \vartheta \mathcal{G} \log \mathcal{G}$.