Should Stock Returns Predictability be "hooked on" Long Horizon Regressions?

Theologos Dergiades

Department of International and European Studies
University of Macedonia, Greece
e-mail: dergiades@uom.edu.gr

Panos K. Pouliasis ‡

Cass Business School
City, University of London, United Kingdom
e-mail: panos.pouliasis@city.ac.uk

Abstract

This paper re-examines stock returns predictability over the business cycle using price-dividend and price-earnings valuation ratios as predictors. Unlike prior studies that habitually implement long-horizon/predictive regressions, we conduct a testing framework in the frequency domain. Predictive regressions support no predictability; in contrast, our results in the frequency domain verify significant predictability at medium and long horizons. To robustify predictability patterns, the analysis is executed repetitively for fixed-length rolling samples of various sizes. Overall, stock returns are predictable for wavelengths higher than five years. This finding is robust and independent of time, window size and predictor.

Keywords: Stock Returns; Long-Horizon Predictability, Frequency Domain

***JEL* Classification**: C10, G17, G32

----

‡ Corresponding author. Cass Business School, City, University of London, 106 Bunhill Row, London EC1Y 8TZ, UK.

"Prediction is very difficult, especially if it's about the future"

Niels Bohr (Nobel laureate in Physics)

1. Introduction

In financial economics among the various financial assets, the body of work exploring aspects of stock returns predictability is voluminous (e.g. Campbell and Hamao, 1992; Boucher, 2007; Henkel *et al.*, 2011; Wen *et al.* 2015; Dergiades *et al.*, 2020). The prevailing view is that stock return variation is predictable, but principally at long horizons. This view has been communicated very eloquently by Cochrane (1999) as one of the most central *New Facts in Finance*. Predictive (or long horizon) regressions constitute methodologically the Chevy Cavalier of this research area (Rapach and Zhou, 2013). In particular, predictability is assessed by regressing stock returns on variables that can act as leading indicators. Nevertheless, for reasons related to the statistical inference emanating from the predictive regressions and the robustness of the conducted pseudo-forecasting exercises, the debate on stock returns predictability is still open (for example see Rapach and Wohar, 2005; Boudoukh *et al.*, 2018).

The current consensus in the existing literature, regarding the derived statistical inference from predictive regressions, could be summarized as follows: First, the Stambaugh (1999) bias that stems in the presence of small samples and persistent predictors, such as earnings-price and dividend-price ratios, leads to substantial size distortions, when testing the null hypothesis of no-predictability (see also Boudoukh *et al.*, 2018). Second, the Berkowitz and Giorgianni (2001) criticism on the inherent inconsistency of the linear predictive regression structure. Rationally, such linear framework should imply *"predictability at all horizons or predictability at no horizon"* (Rapach and Wohar, 2005; p. 328); since, in this context, long horizon forecasts are extrapolations of short horizon forecasts. Third, as Valkanov (2003) argues, the utilization of overlapping observations intensifies the problem of serial-correlation and leads to non-robust standard-errors and misleading inferences; see also Boudoukh *et al.* (2018). For the above reasons, approaches that lead to more robust inferences than predictive regressions have been proposed by several authors, such as Boudoukh *et al.* (2008), among others.

The present paper contributes to literature by aiming an alternative framework to re-examine stock returns predictability. To circumvent the concerns and shortcomings associated with predictive regressions, we shift our analysis to the frequency domain. Econometric methods in the frequency domain are extensively implemented to deal with issues related to financial economics (e.g., Granger, 1969); yet, evidence from such methods is scant for the predictability of stock returns. Exception is Sizova (2014, p. 261), who argues that "*there exist methods, defined in the frequency domain, that provide a superior fit for testing the long-run predictability compared to that provided by long-horizon regressions*".

Across these lines, our testing approach is based on the frequency analysis of causality proposed by Breitung and Candelon (2006). We can thus examine whether stock returns are predictable (via utilizing the information inherent in valuation ratios fluctuations) by considering a different methodological approach. Towards this direction, we postulate a bivariate vector autoregressive (VAR) framework to disentangle short, medium and long run predictability. This is a different frequency domain framework than the one suggested by Sizova (2014); the author implements the local Whittle (1962) estimator and Robinson's (1994) frequency-domain least squares (FDLS). The methodological framework herein demonstrates several advantages over existing methods frequently employed to assess stock returns predictability. For instance, it allows the identification of causal relationships even if the true inter-dependence between two variables is non-linear. Moreover, it permits clear differentiation between short-run and long-run predictability. Last but not least, the test is robust even in the presence of volatility clustering, a common feature of financial time series. Therefore, in certain circumstances, the Breitung and Candelon (2006) test provides an elegant way to deal with potential hidden channels of causality that would otherwise go undetected.

It is well-documented that predictive regressions are sensitive to parameter instability (see Welch and Goyal, 2008; Bansal *et al.* 2004; Kim *et al.* 2005; Lettau and Nieuwerburgh, 2008; and Yin, A., 2019). Rapach and Wohar (2006) find that parameter instability is present when stock returns are predicted through the price-dividend ratio. Davidson and Monticini (2010) find two structural breaks (1959 and 1996) in the long-run relationship between stock prices and dividends. Further, Welch and Goyal (2008) demonstrate that, even in the presence of significant evidence of in-sample predictive ability, on an out-of-sample basis, popular predictors fail to forecast the equity risk premium (for instability on forecasting relationships, see Lettau and Nieuwerburgh; 2008). To avoid misleading inferences due to parameter instability, a common approach is to estimate repetitively the specification of interest on a rolling basis. But even in this case, there is distrust on the credibility of the predictive regression results, because of controversies predominantly associated with the size of the estimation window. Matters of uncertainty occur when: (i) the window-size is selected in an ad-hoc manner and hence, researchers neglect potential predictive patterns that may have been revealed otherwise (Rossi and Inoue, 2012), (ii) researchers present favourable results after trying various specifications and window sizes; that is the so-called data snooping bias (Boudoukh *et al.*, 2008), and (iii) the selected window-size does not accommodate likely parameter instability (Rapach and Zhou, 2013).

To discount for misleading inferences (due to parameter instability, data snooping and ad-hoc window length selection), we follow the general proposal of Rossi and Inoue (2012). Specifically, we test for predictability across multiple window-lengths of fixed size. The choice of the estimation window length has always been a concern for practitioners since model performance is sensitive to the period under investigation. Rossi and Inoue (2012) recommend verification of predictability based on a summary inference, after examining several window lengths (see also Welch and Goyal, 2008; Rapach *et al.*, 2010). An attractive feature of this procedure is the delivery of credible inferences in the presence of structural breaks, data snooping and/or ignorance over the optimal length of the fixed window size. Consequently, we partition our sample into fixed size rolling windows with different lengths; execute the Breitung and Candelon (2006) test for each of the subsamples and evaluate the predictability of stock returns' variations by obtaining summary statistics. Finally, we provide valuable practical insights, by illustrating the evolution of significant predictability across different horizons. Therefore, our findings remain robust to uncertainties linked to predictive regressions.

By using monthly data for the U.S. economy (1871:1 to 2017:9), we re-examine the predictive content of two valuation ratios over the U.S. stock returns, i.e. the price-earnings and the price-dividend ratios, which constitute the most frequently encountered predictors. Most of the empirical results in asset pricing literature corroborate that the price-dividend and price-earnings ratios have limited or no ability in predicting real stock price growth in the short-run (see also Charles *et al.*, 2017), but significant predictive capacity over the long-run; a pattern that has attracted considerable interest by academics and practitioners alike. There is already abundant literature demonstrating the predictive power of valuation ratios on stock returns. For empirical support on the use of the two measures of fundamental value see, *inter alias*, Fama and French (1988), Campbell and Shiller (1988, 1998) and Rapach and Wohar (2005). Nonetheless, recent studies, such as Boudoukh *et al.* (2018), cast doubts on the reliability of such predictor variables, in the context of predictive regressions, concluding that the potential of long horizon return predictability should be considered with scepticism.

A collective view of our findings in the frequency domain indicates that both considered valuation ratios significantly contribute to predicting stock returns; not however in the short-run. The results are surprisingly different compared to the traditional predictive regression framework, even after controlling for robust standard errors and nonlinearities for the data generating process of the predictor variable. The key message is that both the price-dividend and price-earnings ratios evince predictive capacity over stock returns for medium and long horizons; a finding that cannot be revealed when long-horizon regressions are employed. Our results show that stock returns are predictable when using the information content of the price-dividend (price-earnings) ratio for wavelengths higher, on average, than 17 (35) months. The observed difference in the timing at which predictability occurs is consistent with Gopalan and Jayaraman (2012) and Dergiades *et al.* (2020) who claim that investors pay more attention to dividends than earnings in valuing stocks, supporting the supremacy of dividends over earnings and bring to the fore the debate on earnings management, since earnings are prone to information manipulation.

Finally, it is worth noting that this finding has important implications for portfolio allocation decisions. For example, Samuelson (1969) shows that for an investor with power utility who rebalances his portfolio optimally, investment horizon is irrelevant if asset returns are i.i.d. However, variation in expected returns over time can potentially introduce horizon differentiation effects (Merton, 1973). Given the evidence of predictability in returns, a long-horizon investor should allocate his wealth differently from a short-horizon investor (Merton, 1973; Barberis, 2000).

The remainder of this paper is organized as follows. Section 2 presents the data used and Section 3 the methodology. Section 4 shows the results on stock returns' predictability in the frequency domain, compares results to long-horizon regressions and presents robustness checks to eliminate the uncertainty on window selection and parameter instability. Section 5 concludes.

2. Data and preliminary analysis

We use in real terms monthly data of the S&P 500 stock prices (S_t), dividends (D_t) and earnings (E_t) for the U.S. economy, extracted from Shiller's (2005) database.[1] The sample for S_t and D_t spans from January 1871 to September 2017, i.e. almost 146 years, while the sample for E_t extents from January 1871 to June 2017. We calculate the annual stock returns (r_t) as indicated below:

$$r_t = ln[(S_t + \sum_{i=1}^{12} D_{t-1}) / S_{t-12}] \tag{1}$$

[1] Available at: http://www.econ.yale.edu/~shiller/data.htm

where, $\sum_{i=1}^{12} D_{t-1}$ refers to the 12 past monthly periods sum of D_t. Following Rapach and Wohar (2005), the price-dividend ratio (d_t) and the price-earnings ratio (e_t) is calculated by equations (2) and (3), respectively. The time series plot of all constructed variables is shown in Figure 1.

$$d_t = ln[(S_t / \sum_{i=1}^{12} D_{t-1})] \tag{2}$$

$$e_t = ln[(S_t / \sum_{i=1}^{12} E_{t-1})] \tag{3}$$

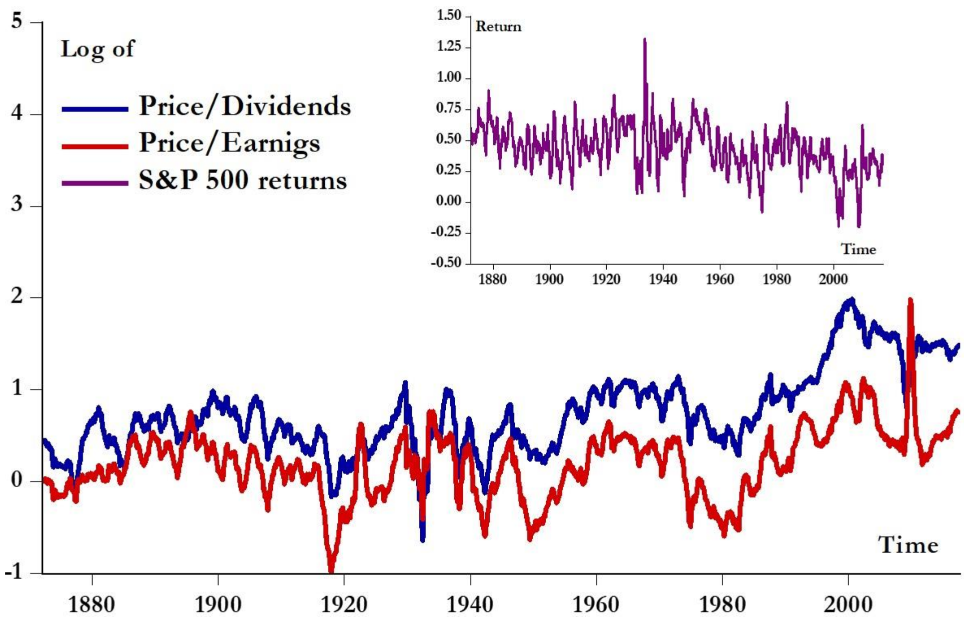


Figure 1. Price-dividend ratio, price-earnings ratio, and S&P 500 stock returns.

Moreover, in Table 1, we report the typical descriptive statistics of the annual stock returns (r_t), the price-dividend ratio (d_t) and the price-earnings ratio (e_t). Finally, as the Breitung and Candelon (2006; henceforth B&C) methodological framework necessitates stationarity for the involved variables, we conduct two commonly used unit-root tests, the augmented Dickey-Fuller (ADF) test and the generalized least squares detrending Dickey-Fuller (GLS-DF) test. The results of both unit-root tests are summarized in Table 2. The conducted inference reveals that all three variables, at the conventional levels of significance, are integrated of order I(0).

Table 1. Descriptive statistics

Variable	Mean	Median	Max.	Min.	Std. Dev.	Skewness	Kurtosis
r_t	0.432	0.435	1.320	-0.200	0.182	-0.197	3.867
d_t	0.742	0.674	1.987	-0.643	0.445	0.579	3.145
e_t	0.192	0.207	1.980	-0.994	0.371	0.115	4.072

Notes: Given the data transformation presented in equations (1) to (3), the sample used to compute the above reported statistics extents from February 1872 to September 2017 (1748 observations) for the variables of the stock returns and the price-dividend ratio, while the respective sample for the price-earnings ratio ranges from February 1872 to June 2017 (1745 observations).

Table 2. Unit root tests

Variable	ADF test		GLS-DF test	
	no trend	trend	no trend	trend
	t-Statistic (k)	t-Statistic (k)	t-Statistic (k)	t-Statistic (k)
r_t	-5.441 (12) ***	-6.458 (12) ***	-3.545 (12) ***	-6.418 (12) ***
d_t	-3.182 (8) **	-4.649 (8) ***	-2.139 (5) **	-4.370 (8) ***
e_t	-5.135 (5) ***	-5.512 (5) ***	-4.490 (5) ***	-5.489 (5) ***

Notes: ADF is the augmented Dickey-Fuller test and GLS-DF is the generalized least squares detrending Dickey-Fuller test. The reported t-Statistic for each test is the test statistic used to conduct the inference for the respective null-hypothesis. For both implemented tests, the lag-length is selected based on the Akaike information criterion. The reported number k in the parentheses is the selected lag-length of the test. Finally, the symbols ***, ** and * denote rejection of the unit root hypothesis at the 0.01, 0.05 and 0.10 significance level, respectively.

3. Econometric methodology

Causality was first put forward by Granger (1969) and has so far been a useful tool to describe relationships in the time domain. A key drawback of causality tests is the restrictive underlying assumption that a single statistical measure is sufficient to essentially explain relationships among variables at all frequencies. In other words, the possibility that causality and feedback mechanisms vary over different frequencies is neglected; this is non-trivial as the choice of data frequency often has significant impact. For a discussion on this, see Lemmens *et al.* (2008) and Narayan and Sharma (2015), among others. Geweke (1982) extends the concept of causality by constructing measures which can be decomposed in time and frequency (see also Hosoya, 1991). In this study, we employ the frequency domain causality test introduced by Breitung and Candelon (2006).

Let f_t be some predictor variable; in our case this is a valuation ratio, i.e. either the price-dividend or the price-earnings ratio. For the stationary $z_t = (r_t \ f_t)'$ two-dimensional vector, we assume a finite-order vector autoregressive (VAR) representation of the form:

$$\Theta(L)\begin{pmatrix} r_t \\ f_t \end{pmatrix} = \begin{pmatrix} \Theta_{11}(L) & \Theta_{12}(L) \\ \Theta_{21}(L) & \Theta_{22}(L) \end{pmatrix}\begin{pmatrix} r_t \\ f_t \end{pmatrix} = \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} \tag{4}$$

Where $\Theta(L) = I - \Theta_1 L - \Theta_2 L^2 - \ldots - \Theta_p L^p$ is a 2×2 lag polynomial of order p with $L^j r_t = r_{t-j}$ and $L^j f_t = f_{t-j}$. For the bivariate white noise process, we define $E(u_t u_t') = V$, as positive and symmetric variance-covariance matrix. Hence, based on the Cholesky decomposition, $H^{-1} = C'C$ with C a lower and C' an upper triangular matrix. This leads to:

$$\begin{pmatrix} r_t \\ f_t \end{pmatrix} = \Psi(L)\begin{pmatrix} \eta_{1t} \\ \eta_{2t} \end{pmatrix} = \begin{pmatrix} \Psi_{11}(L) & \Psi_{12}(L) \\ \Psi_{21}(L) & \Psi_{22}(L) \end{pmatrix}\begin{pmatrix} \eta_{1t} \\ \eta_{2t} \end{pmatrix} \tag{5}$$

Where $\Psi(L) = \Theta(L)^{-1} C^{-1}$ and $(\eta_{1t} \ \eta_{2t})' = C(u_{1t} \ u_{2t}')$, so that $\mathrm{var}(\eta_{1t}) = \mathrm{var}(\eta_{2t}) = 1$ and $\mathrm{cov}(\eta_{1t}, \eta_{2t})$. At frequency ω, it is said that f_t does not Granger cause (no predictive power) r_t if the predictive component of the spectrum of r_t is zero. Hence, causality can be evaluated by:

$$M_{d \to r}(\omega) = ln\left[1 + \frac{\left|\Psi_{12}(e^{-i\omega})\right|^2}{\left|\Psi_{11}(e^{-i\omega})\right|^2}\right] \tag{6}$$

B&C further show that the predictive content of f_t on r_t can be tested through a Fourier transformation on the moving average coefficients (see also Geweke, 1982; Hosoya, 1991). In particular, for predictability at frequency ω, $\left|\Psi_{12}(e^{-i\omega})\right|^2$ should be equal to zero. As $\left|\Psi_{12}(e^{-i\omega})\right|^2$ is a complicated non-linear function, B&C propose to test the same hypothesis using the linear restrictions specified below:

$$\sum_{k=1}^{p} \theta_{12,k} \cos(k\omega) = 0 \quad \text{and} \quad \sum_{k=1}^{p} \theta_{12,k} \sin(k\omega) = 0 \tag{7}$$

Hence, we can test the null hypothesis of no predictability by examining the validity of the above restrictions at frequency $\omega \in (0, \pi)$; that is, by comparing the B&C proposed statistic with the 0.05 critical value from the $\chi^2_{(2)}$ distribution.[2]

4. Empirical results

4.1. Predictability in the Frequency Domain

Table 3 presents the results of the B&C frequency domain causality test for the full sample and a set of selected sub-samples. We first analyze the joint null hypothesis of no predictability for different frequency intervals within the range $(0, \pi)$. Starting with the very short-run, we consider the high frequencies $\omega_i \in (\pi / 3,\ \pi)$ (between 0 and 6 months). Then, the frequencies are partitioned to prespecified monthly intervals, i.e. 6-month increments up to year 4 (48 months) and 12-month increments thereafter up to year 7 (84 months); after which all frequencies are aggregated, i.e. $(84,\ +\infty)$ or $\omega_i < \pi / 41$. To convert the frequency ω_i to time, the transformation $2\pi / \omega_i$ is carried out.

To provide a conservative assessment towards the verification of predictability, we show the maximum *p*-value of the B&C test for each selected frequency range in Table 3. A low *p*-value (lower than the commonly implemented significance levels) implies rejection of the null hypothesis within the corresponding frequency interval. The second column of Panel A reports the full sample results using the price-dividends ratio as a predictor. The maximum *p*-value for each sub-range of frequencies indicates that the null hypothesis cannot be rejected at horizons lower than a year. For all other horizons, predictability is verified at the 0.01 significance level. Similar is the inference when the price-earnings ratio is used as predictor (Table 3, Panel B).

[2] Note that, Breitung and Schreiber (2018), propose a framework where the non-causal hypothesis can be tested for a predefined frequency range instead of single frequencies.

Table 3. Frequency Domain Causality Test (max *p*-values)

Frequency (in months)	Sample						
	1872-2017	1922-2017	1952-2017	1962-2017	1972-2017	1982-2017	1992-2017
	Panel A: Price-dividend ratio (d_t)						
(0,6]	0.953+	0.999+	0.698+	0.669+	0.612+	0.694+	0.529+
(6,12]	0.789+	0.698+	0.191+	0.142+	0.255+	0.639+	0.399+
(12,18]	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.001**	**0.005**
(18,24]	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.002**
(24,30]	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.002**
(30,36]	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.002**
(36,42]	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.001**
(42,48]	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.001**
(48,60]	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
(60,72]	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
(72,84]	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
(84,+∞)	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
	Panel B: Price-earnings ratio (e_t)						
(0,6]	0.993+	0.991+	0.992+	0.999+	0.965+	0.965+	0.961+
(6,12]	0.575+	0.471+	0.434+	0.349+	0.520+	0.578+	0.820+
(12,18]	**0.000**	**0.005**	**0.007**	**0.015**	**0.045**	0.119	**0.098**
(18,24]	**0.000**	**0.001**	**0.001**	**0.006**	**0.017**	0.111	**0.025**
(24,30]	**0.000**	**0.000**	**0.001**	**0.007**	**0.018**	**0.085**	**0.026**
(30,36]	**0.000**	**0.000**	**0.001**	**0.006**	**0.017**	**0.039**	**0.027**
(36,42]	**0.000**	**0.000**	**0.000**	**0.002**	**0.099**	**0.009**	**0.027**
(42,48]	**0.000**	**0.000**	**0.000**	**0.000**	**0.002**	**0.005**	**0.026**
(48,60]	**0.000**	**0.000**	**0.000**	**0.000**	**0.001**	**0.002**	**0.025**
(60,72]	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.001**	**0.024**
(72,84]	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.001**	**0.024**
(84,+∞)	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.001**	**0.023**

Notes: The table illustrates for different samples the maximum *p*-values after partitioning the full frequency range (0, π) to shorter sub-ranges which are mapped to the time domain. That is, the frequency sub-range [π/3, π) corresponds to a time horizon that spans between 0 and 6 months (0,6]. The **dark blue** values imply predictability at the 0.01 significance level. The **light blue** values indicate predictability at the 0.05 significance level. The **turquoise** values signify predictability at the 0.1 significance level, while the **red** values suggest no predictability at all. Numbers in bold signify predictability at the conventional levels of significance. Finally, the superscript + indicates that the average of the *p*-values (instead of the maximum) is greater than the conventional levels of significance.

For robustness, we also split the sample into sub-periods. Hence, the remaining columns of Table 3 contain the testing results for six sub-samples (1922-2017, 1952-2017, 1962-2017, 1972-2017, 1982-2017 and 1992-2017). For the price-dividends ratio (Table 3, Panel A), the results reveal identical predictive pattern as in the full sample; that is, no differentiation can be identified across the examined sub-samples. Looking at the price-earnings ratio (Table 3, Panel B), the overall inference is qualitatively like the price-dividends ratio. In terms of predictive horizons and level of significance, for the first three sub-samples (1922-2017, 1952-2017, 1962-2017) the revealed predictive pattern is undistinguishable to the respective pattern of the full sample. For the last three sub-samples, while predictability can be justified approximately for the same time horizons, this occurs mainly at higher levels of significance (e.g. 0.05 or 0.1). In summary, the B&C testing procedure implies strong predictability patterns in the medium and long run.

4.2. Comparison with Predictive Regressions

Traditionally, stock return predictability is assessed by regressing equity returns on variables that can potentially explain future movements in stock prices; see Fama and French (1988) and Valkanov (2003) among others. A subsequent step is to contrast the findings of the previous section with estimates derived by standard predictive regressions, using the same dataset for the same horizons. The adopted specification is presented below:

$$r_{t+h}^{h} = a + b_h f_t + u_{t+h}^{h} \tag{8}$$

Table 4 reports the OLS estimates of b_h for the full sample and the corresponding *t-statistics* for each predictor. Given the overlapping structure for the observations of r_{t+h}^{h}, standard errors are corrected for heteroscedasticity and serial correlation, using the Newey and West (1987) adjustment based on the Bartlett kernel. Like the previous section, the examined horizons span from 1 to 84 months. Contrary to the results of the frequency domain causality test, the *t-statistics* in Table 4 fail to verify significant predictability for either valuation ratio for all horizons.

For robustness, we also resort to bootstrap simulation methods; see, among others, Kilian (1999) and Rapach and Wohar (2005). In this setting, we calculate bootstrapped *p*-values under the following procedure. Initially, we estimate the data generating process for the returns of stock prices under the null hypothesis of no predictability based on the actual data. We then assume a first order autoregressive (AR) process for the two valuation ratios of interest. Having at our disposal the above estimates, we construct pseudo samples for the stock returns and the two valuation ratios by randomly drawing with replacement from the corresponding residuals, i.e. random walk for r_t and AR(1) for d_t or e_t. To preserve the contemporaneous correlation structure, residuals are drawn in tandem for both processes. Next, for the horizons of interest, we re-estimate the predictive regression of equation (8), storing the Newey and West (1987) adjusted *t-statistics*. The above procedure is repeated 1000 times to generate an empirical distribution of *t-statistics* for each horizon under the null hypothesis of no predictability. For more technical details, the reader is referred to Rapach and Wohar (2005). As inferred from Table 4, the reported bootstrapped *p*-values (in brackets) confirm the findings so far, i.e. the null hypothesis of no predictability cannot be rejected at the conventional significance levels.

Table 4. Predictive regression estimates

Horizon h (in months)	Price-dividend ratio				Price-earnings ratio			
	$\hat{b}_h$	*t*-Stat.	*p*-values		$\hat{b}_h$	*t*-Stat.	*p*-values	
1	0.001	-0.178	(0.770)	[**0.836**]	0.002	0.480	(0.772)	[**0.846**]
6	-0.010	-0.619	(0.566)	[**0.586**]	-0.008	-0.413	(0.472)	[**0.568**]
12	-0.036	-1.231	(0.310)	[**0.356**]	-0.036	-0.916	(0.316)	[**0.426**]
18	-0.059	-1.382	(0.309)	[**0.344**]	-0.061	-1.145	(0.264)	[**0.334**]
24	-0.074	-1.433	(0.264)	[**0.328**]	-0.079	-1.245	(0.216)	[**0.330**]
30	-0.083	-1.431	(0.300)	[**0.332**]	-0.086	-1.190	(0.232)	[**0.316**]
36	-0.099	-1.481	(0.302)	[**0.304**]	-0.093	-1.109	(0.328)	[**0.360**]
42	-0.124	-1.599	(0.300)	[**0.286**]	-0.106	-1.086	(0.304)	[**0.406**]
48	-0.149	-1.736	(0.236)	[**0.282**]	-0.121	-1.133	(0.308)	[**0.392**]
60	-0.192	-1.988	(0.190)	[**0.294**]	-0.168	-1.412	(0.282)	[**0.330**]
72	-0.205	-1.989	(0.252)	[**0.291**]	-0.193	-1.490	(0.246)	[**0.332**]
84	-0.230	-2.077	(0.272)	[**0.288**]	-0.239	-1.646	(0.208)	[**0.326**]

Notes. The table presents for different horizons a) the coefficient estimates, b) the corresponding *t-statistics* (after correcting the standard errors for heteroscedasticity and serial correlation), c) the bootstrapped *p*-values in brackets (·) under the assumption of a linear data-generating process and d) the bootstrapped *p*-values in square brackets [·] under the assumption of a nonlinear data-generating process (for this reason an Exponential Smooth Transition Autoregressive (ESTAR) model is employed). Finally, the **light red** *t-statistics*, the **dark red** *p*-values (under a linear DGP) and the **bold dark red** *p*-values (under a nonlinear DGP) signify no predictability at the conventional levels of significance (e.g. 0.01, 0.05 and 0.1).

The pattern evident in Table 4 is in line with the view that under a linear setting, it is not justified to observe diverse patterns of predictability over different horizons. For example, Berkowitz and Giorgianni (2001) argue that the assumption of linearity implies either predictability or no predictability, irrespective of the horizon. The predictive regression results thus draw a very different picture of predictability compared to the evidence presented in Table 3. Considering this, to further validate our inferences, we relax the assumption of a linear DGP for the two valuation ratios considered. That is, instead of an AR process we adopt the Exponential Smooth-Transition AR (ESTAR) modelling specification, proposed by Kilian and Taylor (2003); for a similar application with annual data, see Rapach and Wohar (2005). Table 4 shows that the tabulated bootstrapped *p*-values (in square brackets) retain the same significance pattern under the assumption of a nonlinear DGP, i.e., the tests preclude predictability at all horizons.

The predictive regression results with monthly data fail therefore to find significance for either valuation ratio (d_t and e_t) at the conventional significance levels. Whilst this is consistent with the results of Boudoukh *et al.* (2008) using annual data, our findings contrast the inference of Campbell and Shiller (1998) and Rapach and Wohar (2005). Both studies use annual data and find that valuation ratios are useful for forecasting stock price changes at long horizons. For example, Rapach and Wohar (2005) confirm the strong ability of valuation ratios to forecast real stock price growth after the fifth year and above.

Finally, the results in Tables 3 and 4 construe mixed assessments. The inference derived from the causality tests in the frequency domain supports the predictability of stock returns. It can be argued that methods that match the frequency spectrum of the data may enclose valuable information content, as causal influences are not uniform across the short-, medium- and long-term horizons. In contrast, when judging predictability solely by predictive regressions conducted in the time domain, part of this information is inevitably lost (Sizova, 2014). Taken together, successful implementation of frequency domain causality can minimize the risk of distorted inference and overcome the limitations associated with the predictive regression framework. Hence, under certain conditions, the B&C test may expose hidden channels of predictability that otherwise would go undetected.

4.3. Robustness on the Predictability Patterns

This section builds on the results derived by the frequency domain testing approach, with purpose to robustify the identified predictive pattern for both employed valuation ratios. Acting so is more than imperative, as substantial changes in the market dynamics are expected to have occurred, given that the start of the sample is dated back to 1871.[3] When dealing with such an extended sample, concern is the presence of structural breaks and their impact on the estimation process. Hence, in the context of Rossi and Inoue (2012), we repetitively test for predictability by adopting a testing framework on a rolling basis over a wide range of windows lengths. This way, we discount the uncertainty over the identified full sample predictability patterns in the presence of breaks.

To this end, the ensuing experiments are conducted using a rolling window estimation scheme as follows. We start with the shortest fixed window-length considered, covering the period from 1871:1 to 1913:8 (500 monthly observations or ≈42 years). This window is used to estimate the parameters of the bivariate VAR specification using annualized r_t and one valuation ratio each time (i.e. either d_t or e_t). To ensure that the estimated specification is kept parsimonious, the optimal VAR lag-length is selected each time based on the AIC criterion. Once the estimation is conducted, we then perform the B&C test. Next, we reconduct the same steps for the subsequent rolling sample, which is constructed by removing the first observation of the previous sample and adding one extra observation at the end of the sample (i.e. 1871:2 to 1913:9). The testing is repeated until the full sample of observations is exhausted; in this case, the final sample is 1976:2 to 2017:9 (in total the B&C test is conducted 1250 times). The above described testing process is repeated for a set of fixed-length rolling windows comprising of 600, 700, 800, 900, 1000, 1100 and 1200 observations. Using a range of different window sizes minimizes the likelihood of our findings being subjected to a window size selection bias. This way, the results offer insights on the dynamic patterns and intensity of stock return predictability throughout time.

[3] These changes may be attributed to several coexisting reasons such as the market depth, market liquidity, number of firms, trading system etc.

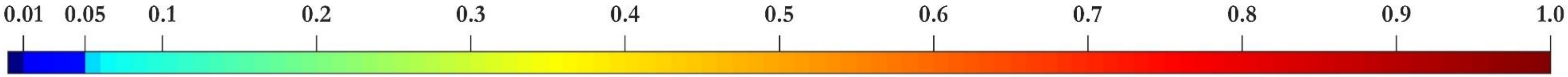
The B&C testing results for d_t and e_t, using rolling windows with fixed size of 500 and 1000 observations, are presented in Figures 2.a and 3.a, which illustrate the distribution of the estimated *p*-values in a three-dimensional coordinate system.[4] The vertical axis depicts the magnitude of the *p*-values; the bottom horizontal left-axis displays the end date of every rolling sample, and the bottom horizontal right-axis depicts the frequency at which the *p*-values refer to. The produced surface is colored according to the magnitude of the reported *p*-values (see the color bar above Figure 2.a). Shades of blue imply rejection of the null-hypothesis of no-predictability at α=0.01 and α=0.05, while all other color shades signify the reverse. Figures 2.b and 3.b portray, for each of the 3D figures, the respective contour plot in the two-dimensional space. The time-frequency plane of the contour plot toggles tints according to the magnitude of each *p*-value, following the same coloring pattern as in the 3D Figures.[5]

Concentrating on the first set of rolling windows with 500 observations (Figures 2.a and 2.b), we estimate for the frequency range $(0,\pi)$, 311,250 *p*-values from 1250 rolling samples (first sample 1872:1 to 1913:8, and last sample 1976:2 to 2017:9). At low frequencies, for *a*=0.01, the hypothesis of no-predictability is consistently rejected across almost all samples. As we progressively move to higher frequencies, inference is again robust but reverse, meaning that the null hypothesis is not rejected for most samples. The price-dividends ratio exhibits predictive power towards stock returns only for long and medium-run horizons, while not in the short-run. For all rolling samples jointly, the null-hypothesis is rejected at *a*=0.01 when $\omega_i \in (0,\ 0.17)$, implying that predictability is verified for wavelengths higher than 36 months; note that we focus on whether predictability is verified jointly at the 100% of the implemented rolling samples.

---

[4] For brevity, we present only the results based on the rolling windows with fixed size of 500 and 1000 observations. The respective figures for rolling windows with fixed size of 600, 700, 800, 900, 1100 and 1200 observations are available upon request.

[5] The contour-plot spots clusters of *p*-values not visible in the 3D figure (e.g. troughs surrounded by peaks).



Rolling-window with 500 obs.

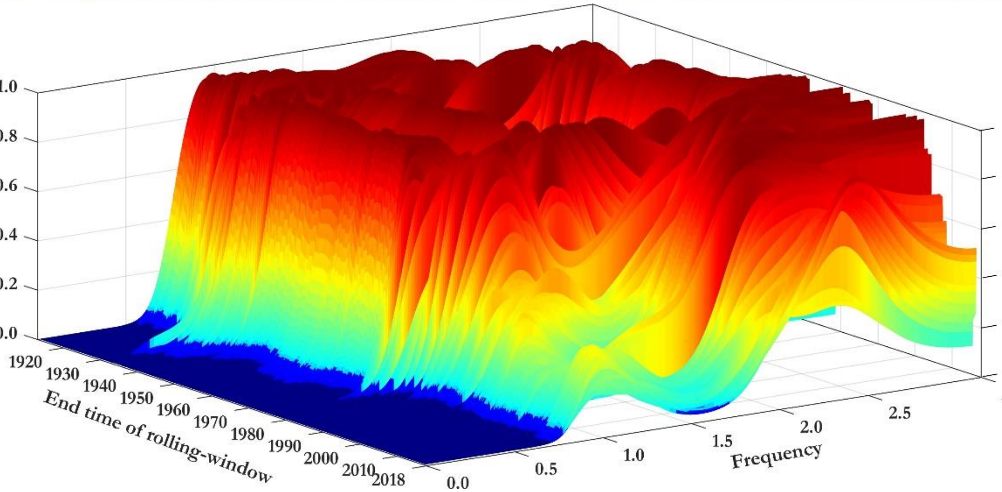


Figure 2.a. *p*-values distribution in a 3D coordinate system (price dividends ratio)

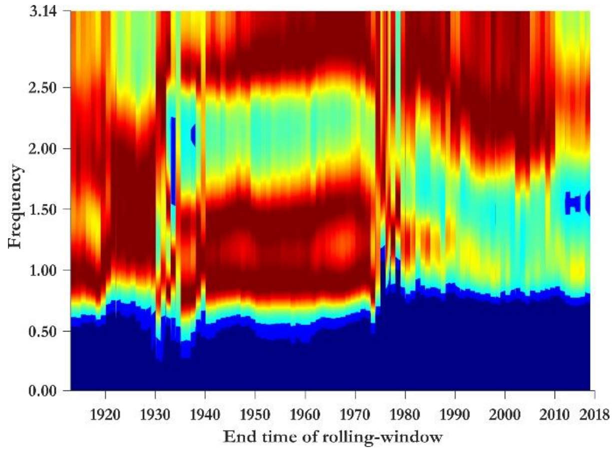


Figure 2.b. *p*-values contour plot

Rolling-window with 1000 obs.

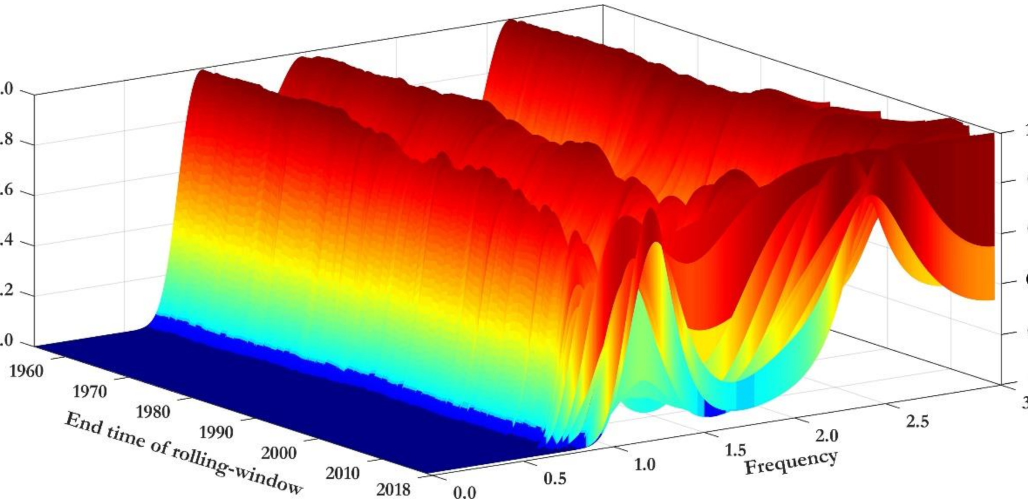


Figure 3.a. *p*-values distribution in a 3D coordinate system (price dividends ratio)

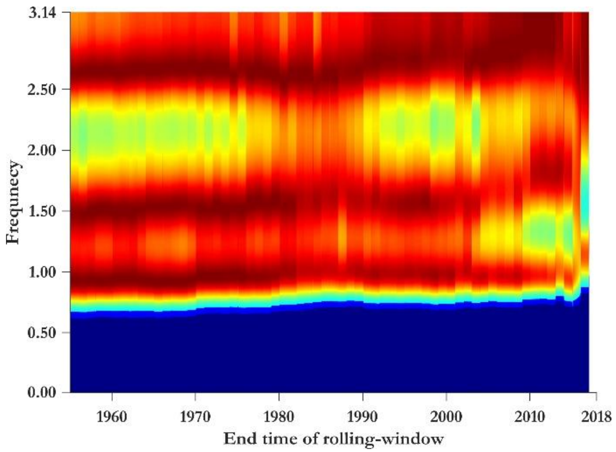


Figure 3.b. *p*-values contour plot

In addition, post-World War II and notably after the 60's, predictability becomes more and more evident at slightly higher frequencies (contour plot of Figure 2.b). In particular, for a = 0.01 there is weaker evidence of predictability for wavelengths between 12 and 36 months $\omega_i \in [0.17,\ 0.53]$. Therefore, from the two plots (Figures 2.a and 2.b), we can deduce that stock return predictability is not constant throughout time but depends on the idiosyncratic characteristics of the underlying period. Hence, there might be significant time-varying patterns, a finding that is consistent with various studies, such as Chen (2009) and more recently Devpura *et al.* (2018), among others.

For example, for the 500-window length, considering the rolling samples that end within the 70's or later, i.e. first sample spanning from 1928:6 to 1970:1 (573 samples in total), the resulting inference indicates that the null-hypothesis (for all rolling samples jointly) is rejected at a=0.01 when $\omega_i \in (0,\ 0.41]$. That is, for this set of samples, predictability can be verified for wavelengths of 15 months or more. Instead, when we account for rolling samples that start from the 70's or later, i.e. beginning sample from 1970:1 to 2011:8 (74 in total; not overlapping with the set of samples mentioned previously), the time-varying nature of stock return predictability is confirmed. The null-hypothesis is now rejected (for all rolling samples jointly) at a = 0.01 for $\omega_i \in (0,\ 0.69]$, implying that predictability is verified for wavelengths of less than a year, i.e. 9 months or more. Overall, these findings provide robust evidence of significant time-varying patterns in the predictability of stock returns.

Figures 3.a and 3.b present the results for the windows with length equal to 1000 observations. For this window size, we estimate in total more than 0.37 million p-values from 750 rolling samples. Based on the distribution of the estimated p-values, a qualitatively equivalent inference as for Figures 2.a and 2.b is deduced. Namely, p-values for low-sized frequencies steadfastly corroborate stock returns predictability, with the respective p-values at high-frequencies showing the opposite. Overall, for the lower length window (500 obs.), the observed variability (across the rolling samples) in the frequency level at which predictability is verified (a = 0.01), reduces as the window length increases. This is expected, provided that small length rolling samples are more sensitive to structural breaks (Rossi and Inoue, 2012); as opposed to large length samples that smooth out the effects of breaks.

Table 5 summarizes the B&C testing results for all examined fixed size rolling windows, that is 500, 600, 700, 800, 900, 1000, 1100 and 1200. From this set of fixed size rolling windows, 2.84 million p-values are estimated from 7200 rolling samples. Panel A in Table 5 illustrates descriptive information for each window size, while Panel B reports the maximum p-value (among all rolling samples of the same size) for selected segments of the frequency range $(0, \pi)$, which are further mapped to the time domain. Hence, for the rolling samples of 500 observations, we may infer (using a=0.01) that on aggregate, stock returns are predictable at wavelengths that are approximately longer than 3 years. For all rolling samples jointly and for each window-length, the intervals of ω, at which the no predictability hypothesis is rejected for a=0.01, range between $(0,\ 0.17]$ and $(0,\ 0.67]$. A collective view of the results suggests that stock returns are predictable for wavelengths higher on average than 17 months. Hence, predictability is evidenced only for medium and long-run horizons.

Next, the same testing strategy is conducted for the price-earnings ratio. The B&C test is executed on a rolling basis for eight different windows sizes, starting from 500 to 1200 observations. The estimated p-values for two window sizes, 500 and 1000, are illustrated in Figures 4.a, 4.b, 5.a and 5.b, respectively, while a summary of results for the remaining window sizes is provided in Table 6. Overall, the resulting inference does not differ qualitatively from the case of the price-dividend ratio. The no predictability hypothesis for the stock returns is primarily rejected at low frequencies, while the opposite is true at high frequencies. We also observe a relative reduction in the variability (across the rolling samples) of the frequency level at which predictability is verified (a=0.01) as the window length progressively increases. The results again indicate that predictability occurs for medium and long-run horizons.

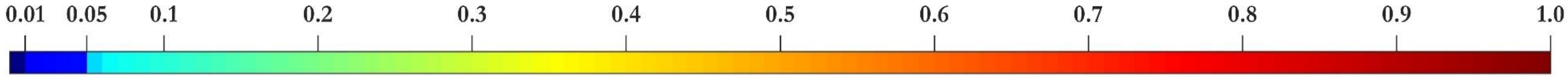
Table 5. Rolling sample summary results for the price-dividend ratio

Panel A								
Summary information	Window length (# of observations)							
	500	600	700	800	900	1000	1100	1200
Window length (in years)	41.7	50.0	58.3	66.7	75.0	83.3	91.7	100.0
# of rolling samples (RS)	1250	1150	1050	950	850	750	650	550
Millions of p-values	0.31	0.34	0.37	0.38	0.38	0.37	0.36	0.33
Joint null rejection by all RS‡	(0, 0.17]	(0, 0.21]	(0, 0.34]	(0, 0.53]	(0, 0.57]	(0, 0.60]	(0, 0.63]	(0, 0.67]
Greater than (in years)	{3.08}	{2.49}	{1.54}	{0.99}	{0.92}	{0.87}	{0.83}	{0.78}

Panel B								
Frequency in months	p-values within horizon intervals (in months)							
(0,6]	1.000⁺	1.000⁺	1.000⁺	1.000⁺	1.000⁺	1.000⁺	1.000⁺	1.000⁺
(6,12]	1.000⁺	1.000⁺	1.000⁺	1.000⁺	1.000⁺	1.000⁺	1.000⁺	0.999⁺
(12,18]	0.188⁺	0.164⁺	**0.060**	**0.007**	**0.002**	**0.001**	**0.000**	**0.000**
(18,24]	**0.069**	**0.042**	**0.011**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
(24,30]	**0.044**	**0.024**	**0.004**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
(30,36]	**0.021**	**0.007**	**0.001**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
(36,42]	**0.007**	**0.002**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
(42,48]	**0.003**	**0.001**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
(48,60]	**0.002**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
(60,72]	**0.002**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
(72,84]	**0.002**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
(84,+∞)	**0.002**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**

Notes: The table reports information about the different Rolling Samples (RS) of fixed length examined. a) ‡ the joint null rejection by all RS, refers to the interval (·] of the ω's in which the null hypothesis of no predictability is rejected jointly by all RS at the 0.01 significance level. b) The symbol # stands for number. c) The reported value in {·} is the conversion of the higher frequency ω in (·], to years (a value of 3.08 implies that predictability is verified at horizons equal or above 3.08 years). Panel B reports, for a specific time range (in months), the maximum p-value among the total number of p-values that are derived from all RS for a given window-length. The **dark blue** values imply predictability at the 0.01 significance level. The **light blue** values indicate predictability at the 0.05 significance level. The **turquoise** values signify predictability at the 0.1 significance level, while the **red** values suggest no predictability at all. Numbers in bold signify predictability at the conventional levels of significance. Finally, the superscript ⁺ indicates that the average of the p-values (instead of the maximum) is greater than the conventional levels of significance.

Nevertheless, even though our findings remain robust, we may note a differentiation in the identified predictability pattern of stock returns which is naturally attributed to the choice of predictor (price-dividends ratio or price-earnings ratio). In the case of the price-earnings ratio, focusing jointly on all rolling samples of 500 observations (see Figure 4.a), the null hypothesis is rejected (a=0.01) when $\omega_i \in (0,\ 0.11]$, implying predictability for wavelengths of more than 57 months. The time period at which predictability exists is higher by 21 months compared to the respective predictive ability of the price-dividends ratio (see Figure 2.a). What is more, based on all rolling samples jointly for every window size, the intervals of ω at which the null hypothesis is rejected (a=0.01) range between $(0,\ 0.11]$ and $(0,\ 0.38]$ (see Table 6). Thus, stock returns are



Rolling-window with 500 obs.

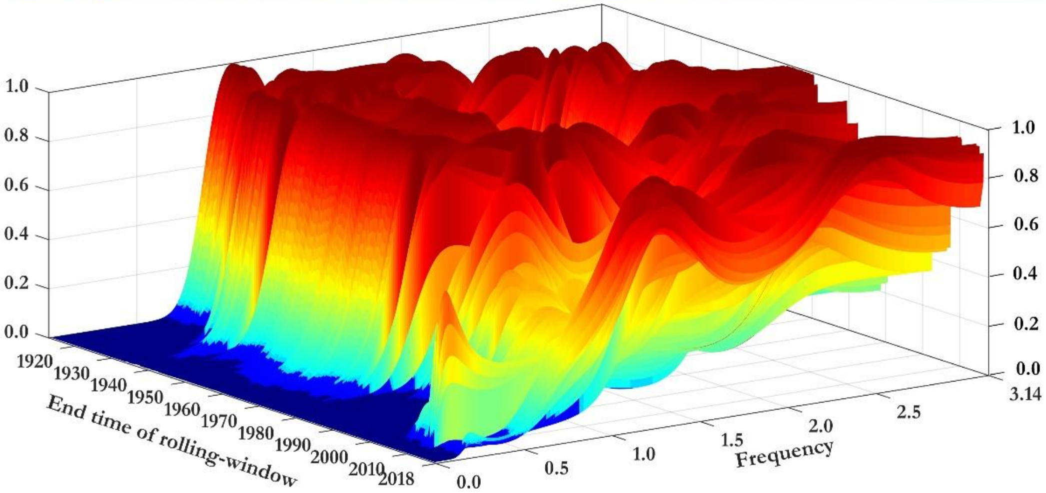


Figure 4.a. *p*-values distribution in a 3D coordinate system (price earnings ratio)

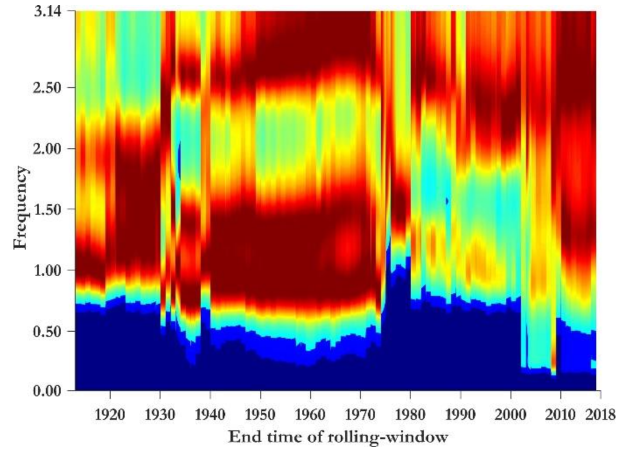


Figure 4.b. *p*-values contour plot

Rolling-window with 1000 obs.

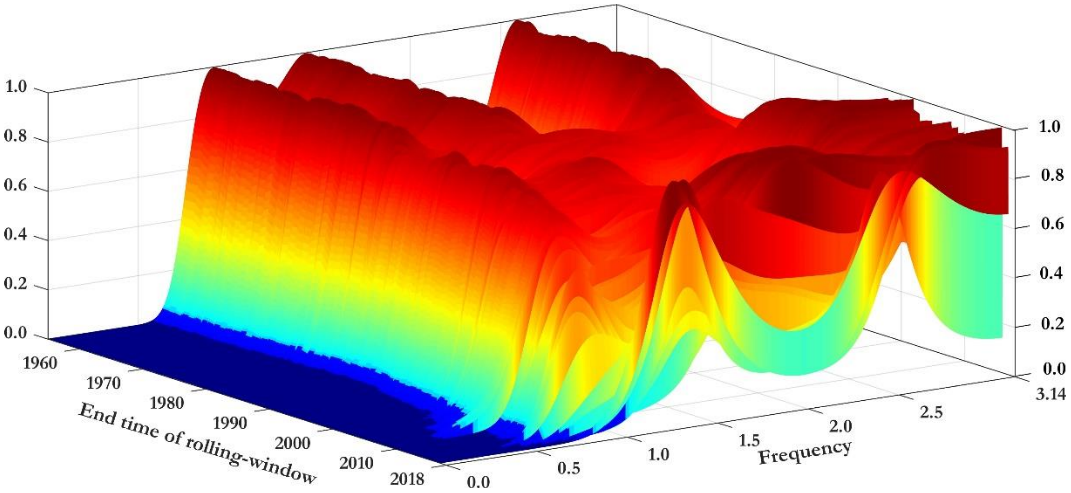


Figure 5.a. *p*-values distribution in a 3D coordinate system (price earnings ratio)

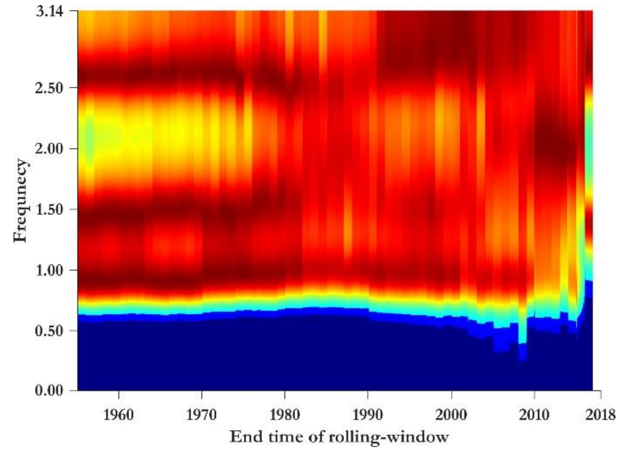


Figure 5.b. *p*-values contour plot

predictable on average at wavelengths higher than 35 months, i.e. 18 months more than the respective mean for the price-dividends ratio. To this end, the price-dividends ratio offers predictability at shorter horizons compared to the price-earnings ratio. This difference is attributed to the quality of the information content conveyed by each valuation ratio. Although both valuation ratios, provide valuable signals for pricing stocks, these very signals may be blurred by noise, which may be of a heterogeneous degree in each case.

As earnings are more prone to information manipulation (see Gopalan and Jayaraman, 2012; Aharony and Swary, 1980), a lower signal-to-noise ratio for earnings relative to dividends is expected. Hence, investors will pay more attention to dividends than earnings in valuing stocks as the signal is stronger, and a non-uniform predictive capacity over the stock prices is anticipated. Overall, this empirical finding is consistent with Aharony and Swary (1980) who find that dividends convey information over and above earnings and Dergiades *et al.* (2020), who attribute the observed difference in the timing at which predictability occurs for the two valuation ratios in that investors pay more attention to dividends than earnings in valuing stocks.

Focusing again on the low length windows (e.g. in Figures 2.a and 4.a with 500 observations), the observation of a general increase in the predictive ability over the post-World War II period cannot be supported now, in contrast to the price-dividends ratio. Instead, time-variation seems to be higher for the price-earnings ratio, confirming that alternative predictors might bring about diverse patterns of predictability. However, the finding that predictability occurs for medium and long-run horizons is still preserved and is robust.

Finally, Table 6 summarizes the B&C testing results for all examined fixed length rolling windows, that is 500, 600, 700, 800, 900, 1000, 1100 and 1200. From this set of fixed size rolling windows, 2.84 million *p*-values are estimated from 7176 rolling samples. Panels A and B report the same information as the respective panels in Table 5. From the results of Table 6, we infer that the horizon at which predictability is verified reduces progressively with the increase of the window size (ranging from 4.76 years and above to 1.38 years and above). Overall, for all rolling samples, independent of size, the stock returns predictability at short horizons is rejected.

Table 6. Rolling sample summary results for the price-earnings ratio

Panel A								
Summary information	Window length (# of observations)							
	500	600	700	800	900	1000	1100	1200
Window length (in years)	41.7	50.0	58.3	66.7	75.0	83.3	91.7	100.0
# of rolling samples (RS)	1247	1147	1047	947	847	747	647	547
Millions of *p*-values	0.31	0.34	0.37	0.38	0.38	0.37	0.36	0.33
Joint null rejection by all RS‡	(0, 0.11]	(0, 0.12]	(0, 0.14]	(0, 0.19]	(0, 0.20]	(0, 0.24]	(0, 0.26]	(0, 0.38]
Greater than (in years)	{4.76}	{4.36}	{3.74}	{2.76}	{2.62}	{2.18}	{2.01}	{1.38}
Panel B								
Frequency in months	*p*-values within horizon intervals (in months)							
(0,6]	1.000+	1.000+	1.000+	1.000+	1.000+	1.000+	1.000+	1.000+
(6,12]	1.000+	1.000+	1.000+	1.000+	1.000+	1.000+	0.997+	0.982+
(12,18]	0.439+	0.345+	0.130+	**0.099**	**0.092**	0.129	**0.089**	**0.034**
(18,24]	0.611+	0.413+	0.161+	0.103	**0.047**	**0.050**	**0.032**	**0.008**
(24,30]	0.633+	0.412+	0.157+	**0.080**	**0.043**	**0.015**	**0.009**	**0.002**
(30,36]	0.511+	0.253+	**0.097**	**0.020**	**0.011**	**0.002**	**0.001**	**0.000**
(36,42]	0.205+	0.139	**0.043**	**0.002**	**0.001**	**0.000**	**0.000**	**0.000**
(42,48]	**0.065**	**0.056**	**0.009**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
(48,60]	**0.032**	**0.015**	**0.002**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
(60,72]	**0.008**	**0.001**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
(72,84]	**0.005**	**0.001**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
(84,+∞)	**0.006**	**0.001**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**

Notes: The table reports information about the different Rolling Samples (RS) of fixed length examined. a) ‡ the joint null rejection by all RS, refers to the interval (·] of the ω's in which the null hypothesis of no predictability is rejected jointly by all RS at the 0.01 significance level. b) The symbol # stands for number. c) The reported value in {·} is the conversion of the higher frequency ω in (·], to years (a value of 3.08 implies that predictability is verified at horizons equal or above 3.08 years). Panel B reports, for a specific time range (in months), the maximum *p*-value among the total number of *p*-values that are derived from all RS for a given window-length. The **dark blue** values imply predictability at the 0.01 significance level. The **light blue** values indicate predictability at the 0.05 significance level. The **turquoise** values signify predictability at the 0.1 significance level, while the **red** values suggest no predictability at all. Numbers in bold signify predictability at the conventional levels of significance. Finally, the superscript + indicates that the average of the *p*-values (instead of the maximum) is greater than the conventional levels of significance.

5. Conclusions

This paper revisits the popular issue of stock return predictability using data for the period spanning from 1872 to 2017. In doing so, it contributes to the extant literature via examining the usefulness of frequency domain methods based on valuation ratios for horizons that extent over and above the business-cycle. The wide implementation of long horizon regressions to investigate the above hypothesis, involves complications linked to the robustness of the derived statistical inference and the conducted pseudo-forecasting exercises. Unlike prior studies, we re-examine the hypothesis that stock returns are predictable from a new angle, in the sense that apart from the traditionally implemented predictive regressions, we assess the same hypothesis by working in the frequency domain. To this end, based on the work of Breitung and Candelon (2006), we adopt a frequency domain causality test for exploring predictability at predefined frequencies.

Our results reveal that the statistical inference critically depends on the choice of method. With predictive regressions, there is no evidence of predictability at long horizons, even when we correct for heteroskedasticity and serial correlation and/or after re-assessing the significance levels based on bootstrap simulation methods; as in, *inter alios*, Kilian (1999) and Rapach and Wohar (2005). The same inference persists no matter if we consider a linear or nonlinear structure for the stock returns' data generating process. Interestingly, the frequency domain causality approach reveals strong evidence in support of stock return predictability for medium to long-term horizons.

Within the frequency domain framework, our analysis is executed in a dynamic manner, as the verification of predictability is investigated in a rolling basis. Our tests are conducted repetitively for fixed size rolling windows across different lengths. Our general findings remain robust to the choice of window length. Specifically, to investigate the hypothesis of no predictability we estimate in total 5.68 million p-values from 14376 rolling samples. Moreover, the general predictive inference remains robust to the choice of predictor variable, i.e. price-dividends or price-earnings. Still, the former predictor is more persistent in predictive ability in the short and medium-term horizons. This implies that investors pay more attention to dividends than earnings in valuing stocks. In addition, the price-dividends ratio ability to predict stock returns appears stronger today compared to the pre-World War II era. Overall, the overarching finding of this work is that both price-earnings and price-dividends ratios evince predictive power towards stock returns only for medium- and long-horizons.

Our research postulates that higher levels of predictability are to be expected at longer horizons. This finding has a certain economic significance in the asset pricing literature (see Campbell *et al.*, 1997; Cochrane, 2001). Such empirical evidence forms the basis to formulate theoretical asset pricing models and active portfolio management strategies[6]. Contrary to the results of the frequency domain causality test, the traditional predictive regression framework fails to verify significant predictability, even after controlling for robust standard errors and nonlinearities. Therefore, practitioners should be cautious on the methods employed to identify predictability and assess model selection procedures. The reason is that potential gains to long-term dynamic investment strategies are likely to be hard to detect; when implementing, *inter alia*, tactical asset allocation strategies, market timing and performance evaluation. Instead, this paper provides an appropriate approach based on the frequency domain framework, that allows clear differentiation between short-run and long-run predictability and is robust to volatility clustering and nonlinearities.

Directions for future research include the implementation of the proposed approach to examine predictability of returns using different predictors and financial variables, considering the extensive set of potential stock return predictors. Alternatively, evidence for predictability associated with equity portfolios sorted based on, for example, size or book-to-market, will allow to shed new light on potential differentiation in the power and persistence of predictability across time and frequencies.

[6] For instance, Barberis (2000) examines the effects of asset return predictability on optimal portfolio choice. The author shows that shows that sensible portfolio allocations for short- and long-horizon investors can be very different in the context of predictable returns while predictability remains important in the presence of estimation risk; with the exception of buy-and-hold portfolios with horizons of many years (see also Cochrane, 1999; Stambaugh, 1999).

References

Aharony, J., & Swary, I., (1980). Quarterly dividend and earnings announcements and stockholders' returns: An empirical analysis. *Journal of Finance* 35(1), 1-12.

Bansal, R., Tauchen, G., & Zhou, H. (2004). Regime-shifts, risk premiums in the term structure, and the business cycle. *Journal of Business & Economic Statistics,* 22(4), 396-409.

Barberis, N., (2000). Investing for the long run when returns are predictable. *Journal of Finance*, 55(1), 225-64.

Berkowitz, J., & Giorgianni, L. (2001). Long-horizon exchange rate predictability? *Review of Economics & Statistics*, 83(1), 81-91.

Boucher, C. (2007). Asymmetric adjustment of stock prices to their fundamental value and the predictability of US stock returns. *Economics Letters*, 95(3), 339-347.

Boudoukh, J., Israel, R. & Richardson, M. (2018). Long-horizon predictability: A cautionary tale. *Financial Analysts Journal*, 75(1), 1577-1605.

Boudoukh, J., Richardson, M., & Whitelaw, R.F. (2008). The myth of long-horizon predictability. *Review of Financial Studies*, 21(4), 1-14.

Breitung, J., & Schreiber, S. (2018). Assessing Causality and Delay within a Frequency Band. *Econometrics & Statistics*, 6, 57-73.

Breitung, J., & Candelon, B. (2006). Testing for short- and long-run causality: A frequency-domain approach. *Journal of Econometrics*, 132(2), 363-378.

Campbell, J.Y., & Hamao, Y. (1992). Predictable stock returns in the U.S.A. and Japan: a study of long-term capital market integration. *Journal of Finance*, 47(1), 43-69.

Campbell, J.Y., Lo, A.W., & MacKinlay, C.A., (1997). *The Econometrics of Financial Markets*. Princeton, NJ: Princeton University Press.

Campbell, J.Y., & Shiller R.J. (1988). Stock prices, earnings, and expected dividends. *Journal of Finance,* 43(3), 661-676.

Campbell, J.Y., & Shiller R.J. (1998). Valuation ratios and the long-run stock market outlook. *Journal of Portfolio Management*, 24(2), 11-26.

Charles, A., Darné, O., & Kim, J.H. (2017). International stock return predictability: Evidence from new statistical tests. *International Review of Financial Analysis*, 54, 97-113.

Chen, L. (2009). On the reversal of return and dividend growth predictability: A tale of two periods. *Journal of Financial Economics*, 92(1), 128-151.

Cochrane, J.H. (1999). New facts in finance. *Economic Perspectives: Federal Reserve Bank of Chicago*, 23(3), 36-58.

Cochrane, J.H., (2001). *Asset Pricing.* Princeton, NJ: Princeton University

Davidson, J., & Monticini, A. (2010). Tests for cointegration with structural breaks based on subsamples. *Computational Statistics & Data Analysis*, 54(11), 2498-2511.

Dergiades, T., Milas C. & Panagiotidis T. (2020). A mixed frequency approach for stock returns and valuation ratios, *Economics Letters*, 187, 08861.

Devpura, N., Narayan, P.K., & Sharma, S.S. (2018). Is stock return predictability time-varying? *Journal of International Financial Markets, Institutions & Money*, 52, 152-172.

Fama, E.F., & French, K.R. (1988). Dividend yields and expected stock returns. *Journal of Financial Economics*, 22(1), 3-25.

Geweke, J. (1982). Measurement of linear dependence and feedback between multiple time series. *Journal of the American Statistical Association*, 77(378), 304-313.

Gopalan, R., & Jayaraman, S., (2012). Private control benefits and earnings management: Evidence from insider-controlled firms. *Journal of Accounting Research* 50(1), 117-157.

Granger, C.W.J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37(3), 424-438.

Henkel, S.J., Martin, S.J., & Nadari, F. (2011). Time-varying short-horizon predictability. *Journal of Financial Economics*, 99(3), 560-580.

Hosoya, Y. (1991). The decomposition and measurement of the interdependence between second-order stationary process. *Probability Theory & Related Fields*, 88(4), 429-444.

Kilian, L. (1999). Exchange rates and monetary fundamentals: What do we learn from long-horizon regressions? *Journal of Applied Econometrics*, 14(5), 491-510.

Kilian, L, & Taylor, M.P. (2003). Why is it so difficult to beat the random walk forecast of exchange rates? *Journal of International Economics*, 60(1), 85-107.

Kim, C.J., Morley, J.C., & Nelson, C.R. (2005). The structural break in the equity premium. *Journal of Business & Economic Statistics*, 23(2), 181-191.

Lemmens, A., Croux, C., & Dekimpe, M.G. (2008). Measuring and testing Granger causality over the spectrum: An application to European production expectation surveys. *International Journal of Forecasting*, 24(3), 414-431.

Lettau, M., & Nieuwerburgh, S.V. (2008). Reconciling the return predictability evidence. *Review of Financial Studies*, 21(4), 1607-1652.

Merton, R., (1973). An intertemporal capital asset pricing model. *Econometrica*, 41(5), 867-887.

Narayan, P., & Sharma, S. (2015). Does data frequency matter for the impact of forward premium on spot exchange rate? *International Review of Financial Analysis*, 39, 45-53.

Newey, W.K., & West, K.D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3), 703-708.

Rapach, D.E., Strauss, J.K., & Zhou, G. (2010). Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. *Review of Financial Studies*, 23(2), 821-862.

Rapach, D.E., & Wohar, M.E. (2005). Valuation ratios and long-horizon stock price predictability. *Journal of Applied Econometrics*, 20(3), 327-344.

Rapach, D.E., & Wohar, M.E. (2006). In-sample vs. out-of-sample tests of stock return predictability in the context of data-mining. *Journal of Empirical Finance*, 13(2), 231-247.

Rapach, D.E., Zhou, G. 2013. Forecasting stock returns, in: Elliot, G., Timmermann, A. (Eds.), Handbook of Economic Forecasting, vol. 2. Elsevier, Amsterdam, 329-383.

Robinson, P.M. (1994). Semiparametric analysis of long-memory time series. *Annals of Statistics*, 22(1), 515-539.

Rossi, B., & Inoue, A. (2012). Out-of-sample forecast tests robust to the choice of window size. *Journal of Business & Economic Statistics*, 30(3), 432-453.

Samuelson, P., (1969). Lifetime portfolio selection by dynamic stochastic programming. *Review of Economics and Statistics,* 51(3), 239-246.

Shiller, R.J., 2005. Irrational exuberance. New Jersey: Princeton University Press.

Sizova, N. (2014). A frequency-domain alternative to long-horizon regressions with application to return predictability. *Journal of Empirical Finance*, 28, 261-272.

Stambaugh, R.F. (1999). Predictive regressions. *Journal of Financial Economics*, 54(3), 375-421.

Yin, A. (2019). Out-of-sample equity premium prediction in the presence of structural breaks. *International Review of Financial Analysis*, 65, 101385.

Valkanov, R. (2003). Long-Horizon Regressions: Theoretical Results and Applications. *Journal of Financial Economics*, 68(2), 201-232.

Welch, I., & Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, 21(4), 1455-1508.

Wen, Y.-C., Lin, P.T., Li, B., & Roca, E. (2015). Stock return predictability in South Africa: The role of major developed markets. *Finance Research Letters*, 15, 257-265.

Whittle, P. (1962). Gaussian estimation in stationary time series. *Bulletin of the International Statistical Institute*, 39, 105-129.