

Essays on Trading Strategies and Long Memory

A Dissertation

Presented to the Graduate School
of

University of Exeter

in Candidacy for the Degree of

Doctor of Philosophy

in

ECONOMICS

by

Dooruj Rambaccussing

Dissertation Directors:

George Bulkley

James E.H. Davidson

February 2012

This thesis is available for Library use on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

I certify that all material in this thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.

Signature:

.....

Abstract

Present value based asset pricing models are explored empirically in this thesis. Three contributions are made. First, it is shown that a market timing strategy may be implemented in an excessively volatile market such as the S&P500. The main premise of the strategy is that asset prices may revert to the present value over time. The present value is computed in real-time where the present value variables (future dividends, dividend growth and the discount factor) are forecast from simple models. The strategy works well for monthly data and when dividends are forecast from autoregressive models. The performance of the strategy relies on how discount rates are empirically defined. When discount rates are defined by the rolling and recursive historic average of realized returns, the strategy performs well.

The discount rate and dividend growth can also be derived using a structural approach. Using the Campbell and Shiller log-linearized present value equation, and assuming that expected and realized dividend growth are unit related, a state space model is constructed linking the price-dividend ratio to expected returns and expected dividend growth. The model parameters are estimated from the data and, are used to derive the filtered expected returns and expected dividend growth series. The present value is computed using the filtered series. The trading rule tends to perform worse in this case. Discount rates are again found to be the major determinant of its success. Although the structural approach offers a time series of discount rates which is less volatile, it is on average higher than that of the historical mean model.

The filtered expected returns is a potential predictor of realized returns. The predictive performance of expected returns is compared to that of the price-dividend ratio. It is found that expected returns is not superior to the price-dividend ratio in forecasting returns both in-sample and out-of-sample. The predictive regression included both simple Ordinary Least Squares and Vector Autoregressions.

The second contribution of this thesis is the modeling of expected returns using autoregressive fractionally integrated processes. According to the work of Granger and Joyeux(1980), aggregated series which are derived from utility maximization problems follow a Beta distribution. In the time series literature, it implies that the series may have a fractional order ($I(d)$). Autoregressive fractionally models may have better appeal than models which explicitly posit unit roots or no unit roots. Two models are presented. The first model, which incorporates an ARFIMA(p,d,q) within the present value through the state equations, is found to be highly unstable. Small sample size may be a reason for this finding. The second model involves predicting dividend growth from simple OLS models, and sequentially netting expected returns from the present value model.

Based on the previous finding that expected returns may be a long memory process, the third contribution of this thesis derives a test of long memory based on the asymptotic properties of the variance of aggregated series in the context of the Geweke Porter-Hudak (1982) semiparametric estimator. The test makes use of the fact that pure long memory process will have the same autocorrelation across observations if the observations are drawn at repeated intervals to make a new series. The test is implemented using the Sieve-AR bootstrap which accommodates long range dependence in

stochastic processes. The test is relatively powerful against both linear and nonlinear specifications in large samples.

Contents

1	Introduction	14
2	Exploiting Price Misalignments	19
2.1	Introduction	19
2.2	Related Literature	21
2.3	Model	23
2.3.1	Models for Forecasting Dividends	27
2.3.2	Stochastic Discount Rate and Dividend Growth	30
2.4	Data and Results	35
2.4.1	Data	35
2.4.2	Monthly Frequency	36
2.4.3	Annual Frequency	47
2.5	Conclusion	52
3	Trading Rule Based on Latent Variable Approach	55
3.1	Introduction	55
3.2	Literature Review	57
3.2.1	Trading Rule	57
3.2.2	Present Value	58
3.2.3	Time Varying Expected Returns and Dividend Growth	59
3.3	Methodology	60
3.3.1	Present Value Model	60

3.3.2	State Space Representation	62
3.3.3	Trading Strategy	64
3.4	Data and Results	65
3.4.1	Properties of Expected Returns and Dividend Growth	67
3.4.2	Trading Rule Returns	69
3.4.3	Tests on the Trading Strategy	71
3.4.4	Robustness of the Rule	72
3.5	Conclusion	76
4	Predictive Ability of Expected Returns	78
4.1	Introduction	78
4.2	Literature review	79
4.3	Methodology	81
4.3.1	State Space Representation	81
4.3.2	Predictive Accuracy	83
4.3.3	Stambaugh Bias	85
4.4	Results	86
4.4.1	Statistical Properties of Expected Returns and Growth	87
4.4.2	Predictive Accuracy	88
4.4.3	Out-of-Sample Forecast	90
4.4.4	VAR models	94
4.5	Economic Value of Return Predictability	99
4.6	Conclusion	100
5	Modeling the Persistence in Expected Returns	103
5.1	Introduction	103
5.2	Present Value assuming AR(1)	107
5.2.1	Time Series Properties of Latent and Mea- sured Variables	109

5.3	Present Value assuming an ARFIMA(p, d, q)	113
5.3.1	Autoregressive Approximation to the ARFIMA(p, d, q)	115
5.4	Two Step Present Value Model	121
5.4.1	Time Variation in κ, ρ and C_1	123
5.5	Results	124
5.5.1	Data	124
5.5.2	Optimization of State Space Models	124
5.5.3	Applications	131
5.6	2 Step Model Estimation Results	139
5.7	Conclusion	145
6	Test of Long Memory based on Self-Similarity	147
6.1	Introduction	147
6.2	Long Memory	149
6.3	Aliasing	153
6.4	The test	155
6.5	Bias Correction	157
6.6	Asymptotic distribution of the statistic	160
6.7	Properties under the alternative hypothesis	162
6.8	Monte Carlo Experiments	165
6.9	Application	173
6.9.1	Individual Stock Prices	174
6.9.2	Expected Returns	174
6.10	Conclusion	177
7	Conclusion	179
	References	185

A	Essays in Asset Pricing	196
A.1	Appendix 1	196
A.1.1	Holding Return	197
A.1.2	Graphical Plots of Accumulated Wealth	199
A.1.3	Test of Correlated Means	203
A.1.4	Sweeney X Statistic	205
A.1.5	Switching	208
A.1.6	Tests on Forecasting Accuracy	209
A.1.7	Summary Statistics	213
A.1.8	Simulation Results	214
A.1.9	Rolling and Recursive Tests of Stationarity	218
A.2	Appendix 2	222
A.2.1	The Present Value Model.	222
A.2.2	State Space Model assuming AR(1)	223
A.2.3	Other Statistical results	225
A.2.4	Plot of Trading Rule Returns and Probability Distribution of Relative Prices	228
A.3	Appendix 3	231
A.3.1	Monte Carlo Experiment on Truncation lags	231
A.3.2	Summary Statistics	232
A.3.3	Expected Returns	235
A.3.4	Expected Dividend (Earnings) Growth	239
A.3.5	Correlation over time	241
A.3.6	Monte Carlo Experiment	242
A.3.7	Univariate Models	243
A.3.8	κ_t from 2 stage Model	248
A.3.9	Expected Return	249
A.3.10	Dividend and Earnings growth	252
A.3.11	Plot of Trading Returns	256
A.4	Appendix 4	260
A.4.1	Proof of Proposition 1	260

A.4.2 Proof of Proposition 3 260

List of Tables

2.1	Annual rates of Return.	36
2.2	Difference in Means for Sharpe Ratio.	43
2.3	Sweeney Statistic for Trading Strategy.	44
2.4	Annual rates of Return with transaction costs of 0.5.	45
2.5	Accumulated Wealth for Annual Data	48
3.1	Optimization of Present Value Model.	66
3.2	Summary Statistics for Expected and Realized series.	68
3.3	Cumulated Returns over Horizons.	72
3.4	Test of Correlated Means.	73
3.5	Replacement Sampling	75
4.1	Estimation of State Space Parameters.	87
4.2	In-sample Predictability of Expected Returns.	89
4.3	In-sample Predictability of Price-dividend ratio.	89
4.4	Out of sample Mean Squared Error for Expected Returns.	91
4.5	Out of sample Mean Squared Error for Price-dividend ratio.	91
4.6	Results from VAR model with Realized and Expected Returns: Sample 1900-2008.	94
4.7	Results from VAR model with Realized Returns and Price-dividend ratio: Sample 1900-2008.	95
4.8	Recursive Mean Squared Error for Expected Returns.	97

4.9	Recursive Mean Squared Error for Price-dividend ratio.	98
4.10	Economic Value of Predictability.	100
5.1	Estimation of AR(1) and ARFIMA(1,d,0) Model for Dividend data.	125
5.2	Estimation of AR(1) and ARFIMA(1,d,0) model for Earnings data.	126
5.3	Tests of Time Variation and Persistence.	131
5.4	Goodness of Fit for Returns Equation.	133
5.5	Goodness of Fit for Dividend(Earnings) Growth Equa- tion.	133
5.6	Success Rate of Trading Strategy.	138
5.7	Summary Statistics of Dividends and Earnings Growth.	140
5.8	Summary Statistics of Expected Returns.	140
5.9	Regression Results for AR(1) and ARFIMA (0,d,0) Specification using Dividend Growth.	142
5.10	Regression Results for AR(1) and ARFIMA (0,d,0) Specification using Earnings Growth	143
6.1	Pure Fractional Process.	166
6.2	ARFIMA(1,d,0) Models.	167
6.3	AR(1) Alternatives.	171
6.4	Nonlinear Dynamic Models	172
6.5	The Two-stage Test procedure, with Skip-rate $n = 8$.	173
6.6	Application to Absolute Returns	174
6.7	Application to Expected Returns.	175
A.1	Cumulated Returns (in percentage) over 5 years of Holding.	198
A.2	Tests of Correlated Means.	203
A.3	Z-score for Individual Models.	204
A.4	Sweeney Statistic assuming 5 percent Transaction cost	206
A.5	Sweeney Statistic assuming 1 percent Transaction cost	206

A.6	Sweeney Statistic assuming 2.5 percent Transaction cost	207
A.7	Equity Holding Periods.	208
A.8	Number of switches	209
A.9	Tests of Normality on Forecast Error.	210
A.10	Forecasting Ability test using Hansen’s Superior Ability test.	211
A.11	Descriptive Statistics for Dividends and Log Dividends.	213
A.12	Simulation Experiment	215
A.13	Returns on Holding Positions.	217
A.14	Tests of Stationarity	226
A.15	Correlation of Realized and Expected Dividend Growth.	227
A.16	Correlation of Realized and Expected Returns.	228
A.17	Monte Carlo experiment on Maximum Likelihood estimates of ‘d’ based on Truncated Autoregressive Models	231
A.18	Summary Statistics for Expected and Realized Returns and Expected and Realized Dividend Growth Rates: 1900-2008	232
A.19	Tests of Stationarity on Expected Returns and Expected Dividend Growth Sample 1900-2008.	233
A.20	Correlation between Expected and Realized Returns up to 5 lags.	234
A.21	Correlation between Expected and Realized Dividend Growth up to 5 lags.	235
A.22	Descriptive Statistics and Stationarity tests on the Expected Returns series for Dividend series.	236
A.23	Descriptive Statistics and Stationarity tests on the Expected Returns series for Dividend series.	237
A.24	Application to Absolute Returns	239

A.25 Descriptive Statistics and Stationarity tests on the Expected Earnings Growth rate.	240
A.26 Application to Absolute Returns	241
A.27 Correlation of Expected Returns from ARFIMA(1,d,0) specifications.	242
A.28 Monte Carlo Experiment.	243
A.29 Univariate Regressions: Expected returns using divi- dend growth specification.	244
A.30 Univariate Regressions: Expected returns using earn- ings growth specification	246
A.31 Univariate Regressions on Dividend Growth	247
A.32 Univariate Regressions on Earnings Growth	247

List of Figures

1.1	Price and Present Value of S&P500 index.	15
2.1	Accumulated Returns- Forecast Model 3a and Discount Factor A.	38
2.2	Accumulated Returns- Forecast Model 1a and Discount Factor A.	39
2.3	Accumulated Returns under Buy and Hold, Trading Strategy and Filter Rule.	49
2.4	Holding Positions for Filter Rule, Trading Model and Optimal Profit.	50
2.5	Correlation of Holding Position for Filter Rule and Forecast Model with Ideal Profit.	51
3.1	Plot of Expected and Realized Returns.	67
3.2	Plot of Expected and Realized Dividend Growth Rate.	68
3.3	Cumulated Returns from Trading Strategy.	70
3.4	Plot of Discount Rate.	76
4.1	Plot of Root Mean Squared Error for Expected Returns.	92
4.2	Plot of Root Mean Squared Error for Price-dividend ratio.	92
5.1	Impulse Response for Consumption Growth - AR(1) .	135
5.2	Impulse Response for Consumption Growth - ARFIMA(1,d,0).	135

6.1	ARFIMA (0, 0.4,0)	169
6.2	Bilinear Model	169
6.3	SETAR Model	169
6.4	Markov Switching	170
6.5	ESTAR Model	170
A.1	Cumulated Wealth from January 1901 to December 2007 for Discount Rate A and Forecast Model A	199
A.2	Cumulated Wealth from January 1901 to December 2007 for Discount Rate A and Forecast Model B	199
A.3	Cumulated Wealth from January 1901 to December 2007 for Discount Rate B and Forecast Model A	200
A.4	Cumulated Wealth from January 1901 to December 2007 for Discount Rate B and forecast model B	200
A.5	Cumulated Wealth from January 1901 to December 2007 for Discount Rate C and Forecast Model A	200
A.6	Cumulated Wealth from January 1901 to December 2007 for Discount rate C and Forecast Model B	201
A.7	Cumulated Wealth from January 1901 to December 2007 for Discount rate D and Forecast Model A	201
A.8	Cumulated Wealth from January 1901 to December 2007 for Discount Rate D and Forecast Model B	201
A.9	Recursive ADF T-stat with MAIC Optimal Lag Selection for Price-dividend Ratio	218
A.10	Rolling ADF T-stat with MAIC optimal Lag Selection	219
A.11	Recursive ADF T-stat with MAIC Optimal Lag Selection for the Price-earnings Ratio	220
A.12	Rolling ADF T-stat with MAIC Optimal Lag Selection for Price-earnings Ratio	220
A.13	Plot of Trading Return against Best Return	229
A.14	Probability Distribution of P_t^*/P_t	230
A.15	Plot of κ_t for Dividend Growth from 1926-2008	248

A.16 Plot of Kappa for Earnings Growth for 1926-2008. . .	248
A.17 Plot of Expected Returns for AR(1) and ARFIMA(1,d,0): Sample 1946-2008 using Dividend Data.	249
A.18 Plot of Expected Returns for AR(1) and ARFIMA(1,d,0) and Realized Returns for 1926-2008 using Dividend Data.	250
A.19 Plot of Expected Returns for AR(1) and ARFIMA(1,d,0): Sample 1926-2008 using Earnings Growth.	250
A.20 Plot of Expected Returns for AR(1) and ARFIMA(1,d,0): Sample 1946-2008 using Earnings Data.	251
A.21 Plot of Dividend Growth for AR(1) and ARFIMA(1,d,0): Sample 1946-2008.	252
A.22 Plot of Dividend Growth for AR(1) and ARFIMA(1,d,0): Sample 1926-2008.	253
A.23 Plot of Earnings Growth for AR(1) and ARFIMA(1,d,0): Sample 1926-2008.	253
A.24 Plot of Earnings Growth for AR(1) and ARFIMA(1,d,0): Sample 1946-2008.	254
A.25 Cumulated Returns based on the Price-dividend ra- tio: Sample 1926-2008.	256
A.26 Cumulated Returns under Price-earnings ratio 1926- 2008.	257
A.27 Cumulated Returns for the Price- dividend ratio: Sam- ple 1946-2008	257
A.28 Cumulated Returns based on the Price-earnings ratio for the sample 1926-2008.	258

Acknowledgements

I thank my first supervisor George Bulkley for his interesting discussions and comments which brought a structure to this thesis. I am indebted to my second supervisor, James Davidson who deserves my utmost gratitude and respect for all the help he has provided me since the past three years. I thank him for those wonderful discussions and also for helping me improve my programming skills. I am also thankful to him for finding me enough teaching hours in order to finance my writing up stage. I also thank John Cochrane, Adam Golinski, Ralph Koijen, Joao Madeira, Allan Timmermann and Simone Varotto for constructive advice.

I also thank Louise Armshaw for helping me proofread this thesis.

I thank participants and discussants in the following conferences:

University of Exeter PhD. Finance Seminar 2009

University of Exeter PhD Economics Seminar 2008, 2009

Cournot Doctoral Days 2011

European Financial Management 2011

Royal Economic Society 2012

Chapter 1

Introduction

The ‘Excess Volatility of Share Prices’ is an expression used to describe excessive price fluctuations beyond the present value of dividends. This anomaly was independently detected by Shiller (1981) and LeRoy and Porter (1981). Financial asset prices, in particular equity, tend to fluctuate more than the price of the conventional consumption good. This finding would not be considered an anomaly if only those fluctuations were matched by fluctuations of equal magnitude in the expected future payoffs. According to classical theory, changes in prices of assets are explained by movements in the expected future payoffs. Contrary to theory, prices are highly volatile in practice, whereas dividend payments are smooth over time. This is illustrated in figure 1.1.

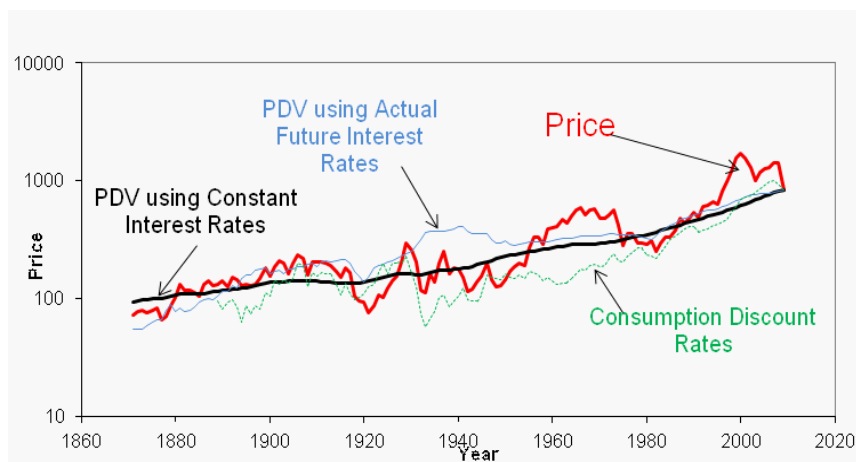


Figure 1.1: **Price and Present Value of S&P500 index.** The figure shows the annual stock price and corresponding present discounted value over time over the period 1871-2009 according to three definitions of the discount rate. (Source Shiller 2011)

Figure 1.1 compares the S&P500 price with the present value over time. The present value is computed using three variants of discount rates. The series abbreviated as ‘PDV using Constant Interest Rates’ refers to the present value when dividends are discounted with a fixed factor¹, which is equal to the sample average of realized returns. In the case of ‘PDV using Actual Future Interest Rates’, the discount factor varies and may be computed as the sum of the real interest rate for the period and the differential between the average realized returns and average real interest rate. ‘Consumption Discounted rates’ discounts real dividends using a rate which is equal to growth in consumption. Computation of the present value shall be examined thoroughly in the next two chapters. In all the three cases, real stock price fluctuates more than the present value. This empirical phenomenon provides a case for the implementation of a trading strategy.

¹The discount factor is constant throughout the sample period. For instance, for $t=1$ and 2, the present value is equal to $\frac{D_1}{r}$ and $\frac{D_2}{r}$ respectively.

Chapter two defines and tests the trading strategy. The premise of the strategy rests in identifying whether markets are under or overpriced with respect to the present value. Based on economic fundamentals, prices may revert to the present value through the actions of arbitrageurs in a volatile market. The trading strategy consists of holding bonds when the market is overpriced and holding equity when the market is underpriced. The strategy is implemented in real-time. Real-time implies that agents decide to go long on bonds or equity based on the information they have at this particular point in time. Computing the present value requires expected values of dividends, dividend growth and the discount rate. Recursive and Rolling forecasts of dividends from three regression schemes are used to proxy expected dividends. The discount rate is computed from a simple mean model and a cointegrating regression framework. When adopted on a monthly basis, the market earns a premium over the Buy and Hold strategy of 4.9% per annum.

Chapter three uses the same underlying theory to identify when equity markets are under or overvalued in real time. A new approach is introduced to derive the discount rate, the dividend growth rate and expected dividends. A structural state space model is constructed using the present value to derive the time series of expected dividend growth and expected returns. These series, typically unobservable in real-time, are filtered based on econometric specifications which show the evolution of these series over time. A battery of tests comparing the rule to the passive Buy and Hold strategy illustrates that the rule performs marginally worse than the Buy and Hold Strategy by 1 % annually.

The two chapters point out that the discount rate computation is an important factor which influences the performance of the strategy. A marginal change in discount rates will lead to considerable

changes in the present value, and hence might tilt the decision of going long on bonds or equity. Simple econometric forecasts work better in the strategy as compared to the present value structural model. The important part played by discount rates in explaining movements in macroeconomic variables is documented by Cochrane (2011).

One of the by-products of chapter three is the decomposition of the price-dividend ratio into expected returns and dividend growth. Chapter four compares the predictability of S&P500 returns from the decomposed expected returns with the price-dividend ratio using both univariate and multivariate models. Expected returns is a potential predictor variable which removes the noise due to the dividend growth in the price-dividend ratio. By filtering out this noise, we may get a better predictor which is both theoretically and statistically motivated. Secondly, the best predictor of a realized value may be its expected counterpart. The results show that expected returns does not improve on the price-dividend ratio as a predictor variable. Evidence of predictability was uncovered only over longer horizons, which is mainly due to the econometric property of persistence.

The major contribution of the fifth chapter is to explore long memory properties in expected returns. In Granger and Joyeux (1980), aggregation of micro random walk processes can be shown to be linear and have a long memory component. An important assumption of the previous two chapters is that expected returns follows an autoregressive process of order one (AR(1)). The empirical evidence showed that expected returns are persistent over time. Hence, it may be possible that expected returns are better specified by an autoregressive fractionally integrated model (ARFIMA(1,d,0)) instead of an AR(1). Two models are put forward to derive the ex-

pected returns series assuming an ARFIMA model. In the first model, a finite state space representation of the present value relationship between the dividend-price ratio, dividend growth and the expected returns is presented. The filtered series is used in three applications: analyzing predictability of returns, the relationship between consumption growth and discount rates and also in the present value strategy. The second model consists of deriving expected returns from a two-stage procedure. In the first stage, dividend growth is forecast using three parametric specifications. In the second stage, the forecast dividend is subtracted from the price-dividend ratio to retrieve a series of expected returns. The ARFIMA(1,d,0) is then fitted to the derived series.

Based on the possibility that expected returns may be long memory, long memory properties are formally tested in expected returns based on the skip sampling procedure. Chapter six illustrates the skip sampling procedure. The null hypothesis is that the series is a pure long memory process. Linear and nonlinear alternatives are considered. The test computes the fractional parameter 'd' using the Geweke Porter-Hudak (1983) procedure, using skip sampled observations. The distribution of the skip sampled 'd' is generated using the Sieve-AR bootstrap which accounts for dependence in the estimation. The test depends on the bandwidth adopted. An application to the absolute log returns of three companies is considered.

Chapter seven concludes.

Chapter 2

Exploiting Price Misalignments

2.1 Introduction

A classic proposition of equilibrium search in asset pricing requires that agents will exploit arbitrage opportunities if they can be identified. Market prices will not be equal to the present value if such opportunities exist. With reference to a typical asset, misalignments between the actual price and the corresponding present value may offer profitable opportunities, which may be arbitrated away as the price reverts to the present value. For instance, if prices are higher than the present value of an asset, it would mean that the asset is overpriced. Over time, there should be a downward reversion towards the fundamental value. On the other hand, if price is lower than the present value, there will be an upward adjustment in future periods. A simple trading strategy in such a case, is to hold the asset when it is underpriced and sell it when it is overpriced.

While the actual price is observable, the empirical problem lies in computing the present value. According to standard asset pricing theory, the price of any asset is the conditional expectations of the

sum of the discounted future payoffs. In the case of equity, the present value is equal to the discounted value of the infinite sum of expected future streams of dividends. A potential problem arises when computing the present value in real-time. This is due to the latent nature of these series. The next paragraph elaborates on this problem.

At time t , expected future dividends is not directly measurable. The discount rate is also unobserved since it is an aggregation of individual discount rates. In the learning literature, when the data generating process for an unobservable or latent variable is unknown, agents are assumed to use econometric models to estimate and forecast the expected value. A similar approach is used in this chapter, where both dividends and expected discount rates are forecast from simple time series models.

The trading strategy was first considered by Bulkley and Tonks (1989 and 1991), where the focus was to derive an implicit test of excess volatility through reversion of prices towards fundamentals. However, both papers may be criticized on the grounds of using a fixed discount factor. Discount rates are time varying. Furthermore, an implicit assumption of their work is that empirically the "future is known". The key innovation of the present work is to adopt the real time framework, by allowing dividends, dividend growth and expected returns to vary over time.

Agents use econometric models to forecast variables filtered on the current information set. Three simple linear models are used to forecast dividends one step ahead in each period. The stochastic discount rate and dividend growth rate are computed using OLS. The forecast variables are then used to derive the rational expectations' present value. Model uncertainty is not tackled in this work,

although it is an active area of research.¹ However, the simplicity of the different forecast models insulates this work from strong issues with model mining.

The real-time strategy is based on the above principle. The construct of the present value implies that the latter is computed at time t , with data available only at that point of time. In the corresponding empirical application, the expected returns and dividend growth is forecast using expanding and rolling windows. Use of such windows is known to reduce parameter uncertainty and estimation risk in large samples. The rule is applied to the S&P 500, where mean reversion and excess volatility have been documented (Shiller and Beltratti (1993), Poterba and Summers (1988)). The present value may be extended to individual stocks. However, in this study, we attempt to see if real-time forecasting of dividends and discount rates is successful for monthly data in a stock index where mean reversion has been established. Applying the strategy to individual stocks is a recommended area of study for individual stocks experience momentum and mean reverting effects. However the contribution of this chapter is methodological.

2.2 Related Literature

The trading rule is built in line with equilibrium search theory in asset pricing. The main gist of the rule is to identify periods when the equity market is mispriced and sequentially decide whether to hold equity or bonds depending on the direction of the mispricing. The simple idea is that if the actual price at time t is higher than the present value, wealth is held in bonds. The objective is to avoid a

¹For an interesting overview of model uncertainty in the context of the present value, see Avramov (2002,2004), Timmermann and Granger (2004), Timmermann (1993,1996), Pesaran and Timmermann(1995,2002) and Rey(2005).

potential capital loss when the price of equity reverts to the present value. During that period, the bond return is earned. On the other hand, if the price of equity is lower than the present value, capital gains may be reaped by holding stocks. In a nutshell, the strategy is a market timing mechanism which advises the agent where to hold wealth (bonds or equity).

The market timing strategy was introduced by Bulkley and Tonks (1989) as a test of weak form efficiency for the UK market. They showed that revision in the parameters of a dividend model may explain excess volatility in prices. This work spurred other areas of investigation in the learning literature about the dividend generating process. The rule was also tested against the Buy and Hold Strategy for the S&P 500 market with the same success as in the UK market (Bulkley and Tonks (1992)). The rule performed better than the simple Buy and Hold Strategy. Bulkley and Taylor (1996) use the present value formulae in a conditional VAR model to derive the theoretical price. The objective was to test whether underpriced portfolios tend to generate higher returns than overpriced portfolios over time.

Unlike the previously mentioned studies, our major contribution lies in applying real-time concept to the rule. Firstly, three econometric models are narrowed as potential data generating processes for dividends. Secondly, the update of the parameters is performed using moving estimation windows. Agents equate their expectations of the different variables (such as the discount factor) to the conditional forecasts. The real-time discount rate is computed using rolling and recursive averages of historic returns and the cointegration approach of Fama and French (2002).

2.3 Model

Standard asset pricing theory states that the price of an asset is equal to the first order conditions from the optimization problem of a two period agent faced with the choice of how much to consume and invest at time t . It follows that the present value at time t should be equal to the expected discounted value of the asset's payoff. The resulting first order condition (referred to as the Euler equation) is known as the fundamental asset pricing equation (2.1)².

$$P_t^* = E_t \left(\beta \frac{u'(C_{t+1})}{u'(C_t)} x_{t+1} \right), \quad (2.1)$$

where P_t^* refers to the price (present value) of the asset, E_t is the expectations operator at time t , $\beta \frac{u'(C_{t+1})}{u'(C_t)}$ is the discount factor or the marginal rate of substitution between consumption from time t to $t + 1$, and x_{t+1} is the payoff at time $t + 1$. β is the subjective discount rate. In the case of the stocks, the payoff of x_{t+1} relates to $P_{t+1} + D_{t+1}$, where P_{t+1} is the next period price (at time $t + 1$) and D_{t+1} is the next period dividend (at time $t + 1$). $\beta \frac{u'(C_{t+1})}{u'(C_t)}$ can be abbreviated as m_{t+1} which is unobserved at time t . Therefore, the price of any asset is given by:

$$P_t^* = E_t(m_{t+1}x_{t+1}). \quad (2.2)$$

m_{t+1} is also known as the stochastic discount factor. In the case of stock, the stochastic discount factor is to the risk-adjusted interest rate (or returns). In the two period model, the price of stock can be written as:

$$P_t^* = E_t \left(\frac{1}{1 + r_{t+1}} (P_{t+1} + D_{t+1}) \right). \quad (2.3)$$

²Cochrane (2005) offers an interesting intuitive insight on how this equation is reached. For more advanced optimization models, refer to Ljungvist and Sargent (2004).

By substituting forward iterated values of P_{t+1} to an infinite number of periods, and assuming the transversality condition $\lim_{j \rightarrow \infty} E_t \left(\frac{1}{1+r_{t+j}} P_{t+j} \right) = 0$, the present value is equal to conditional expectations of the sum of discounted dividends (2.4).

$$P_t^* = E_t \left(\left(\frac{D_{t+1}}{1+r_{t+1}} \right) + \left(\frac{D_{t+2}}{(1+r_{t+1})^2} \right) + \dots + \right) \quad (2.4)$$

$$= E_t \sum_{i=1}^{\infty} \left(\left(\frac{1}{1+r_{t+1}} \right)^i D_{t+i} \right). \quad (2.5)$$

(2.5) can be refined to cases when dividends grow over time. The growth of dividends from time t to $t+1$ is given by:

$$D_{t+1} = (1 + \Delta d_{t+1}) D_t$$

where $\Delta d_{t+1} = \ln\left(\frac{D_{t+1}}{D_t}\right)$, and is known as the dividend growth rate. For the general i steps ahead, the conditional expected dividend is given by:

$$D_{t+i} = (1 + \Delta d_{t+1})^i D_t. \quad (2.6)$$

Substituting (2.6) into (2.5), the present value may be written as:

$$P_t^* = E_t \sum_{i=1}^{\infty} \left(\left(\frac{1 + \Delta d_{t+1}}{1+r_{t+1}} \right)^i D_{t+i} \right). \quad (2.7)$$

If the dividend growth rate is higher than the discount rate, such that $r_{t+1} > \Delta d_{t+1}$, it can be shown that:

$$\sum_{i=1}^{\infty} \left(\frac{1 + \Delta d_{t+1}}{1+r_{t+1}} \right)^i = \frac{1 + \Delta d_{t+1}}{r_{t+1} - \Delta d_{t+1}}. \quad (2.8)$$

Replacing (2.8) into (2.7), the present value is equal to

$$P_t^* = E_t \left(\left(\frac{1 + \Delta d_{t+1}}{r_{t+1} - \Delta d_{t+1}} \right) D_t \right).$$

Since $(1 + \Delta d_{t+1})D_t = D_{t+1}$,

$$P_t^* = E_t \left(\frac{D_{t+1}}{r_{t+1} - \Delta d_{t+1}} \right).$$

Assuming that $Cov_t(D_t, (r_{t+1} - \Delta d_{t+1})) = 0$,

$$P_t^* = \frac{1}{E_t(r_{t+1} - \Delta d_{t+1})} E_t(D_{t+1}). \quad (2.9)$$

(2.9) is used to compute the present value. $E_t(D_{t+1})$, $E_t(r_{t+1})$ and

$E_t(\Delta d_{t+1})$ are forecast individually. The rule posits comparing (2.9) with the price of equity at time t to decide the holding position:

$$\begin{aligned} P_t &> P_t^* : \text{Go Long on Bond Market} \\ P_t &< P_t^* : \text{Go long on Equity Market} \end{aligned}$$

The rule is simple; At the end of every sample date t , the present value is computed and compared with the equity price, and the holding position is determined accordingly. If initially, an agent is in bonds, and the computed present value is less than the equity price, (s)he stays in bonds since it is anticipated that prices will fall. In the next period, if prices are below the present value, (s)he switches into equity. Many variants of the model may be considered, for instance, holding the asset for k periods ahead before shifting the asset.

The empirical challenge of the model is to estimate the variables in (2.9). All the components of the present value are unobserved at time t . The variables are forecast from three econometric mod-

els with moving window estimates. The next few paragraphs shall describe the mechanics of moving windows.

Moving windows can be classified as either rolling or recursive windows.³ Recursive windows allow the information set to grow as a new observation is measured.⁴ It is suited to modelling with parameters which do not vary excessively over time. Rolling windows, on the other hand, use a fixed block of observations (the information set), to estimate the regression models. If the window length of the rolling sample is large and there are no strong variations in parameter estimates, the difference in estimates and forecasts between rolling and recursive windows will be small.

Consider a variable x_t which is observed over a sample of T observations, such that $t = 1, \dots, T$. Recursive windows involve taking a subsample N from T , and estimating the model firstly using the first N observations, and forecasting p steps ahead. In the second round, the first $N + 1$ observations are used in estimation. The estimated parameters are then used to forecast p steps ahead. After each round of estimation, the subsample approaches the full sample T . In the k_{th} round, the estimation sample size is $N + k$. After performing the same procedure k times, the number of forecasts is kp for the full sample. In the rolling window models, the subsample remains fixed (N) over the different rounds. Both the initial and terminal points are increased after each round. In the k_{th} round, the sample of estimation is $[k, N + k]$ where k is the first observation and $N + k$ is the final observation. Under certain conditions; recursive windows should provide more consistent estimates.⁵ Rolling windows are appealing when only the most recent observations matter.

³Interestingly Rolling windows are more common in Finance whereas Recursive windows are popular in empirical Macroeconomics.

⁴In this paper, it implies new information entering the market.

⁵Multiple Breaks are an exception.

However, the variance of the rolling window estimates tends to be high.

In the finance literature, moving windows are used mostly for the purpose of robustness in the presence of structural breaks. In the present context, it is in line with real-time market timing decisions. Agents estimate a model using available data and then forecast one step ahead. They compute the present value at time t and decide whether to hold bonds or equity. In the next period, agents re-estimate the model again and forecast one step ahead and decide to stay in the asset or switch.

2.3.1 Models for Forecasting Dividends

The regression models used to forecast dividends have two functional forms: linear and logarithmic. When the model is forecast with logarithmic values, the conditional forecast of the logarithm dividends is exponentiated so that expected dividends are in level forms again.⁶ In the case of monthly data, seasonality is accounted for by including month dummy variables in each model. Relying on seasonal dummies may have its limitations; however, the benefits of applying the rule to a monthly frequency exceed the effects of the marginal seasonality estimation errors. The three models that are used to forecast dividends are as follows:

⁶It is important to note that when log dividends are used, the corresponding forecast is biased, since the expected value of a nonlinear function is not the same as the nonlinear function of the expected value. This is known as Jensen's inequality $E(f(x)) \neq f(E(x))$. Hence $e^{(\log D_{t+1})}$ is a biased estimate of D_{t+1} . One solution to get rid of this problem is to use a Taylor expansion of $e^{(\log D_{t+1})}$ and adjust forecasts with higher order derivatives. While bias may be achieved, it may inflate the variance. In our case, no correction has been made.

Model 1a

$$D_{t+1} = \beta_0(t+1) + \sum_{i=1}^{12} \beta_i I_i + \sum_{i=0}^p \delta_i D_{t-i} + \gamma_1 \left(\frac{D}{P}\right)_{t-1} + \gamma_2 \left(\frac{E}{P}\right)_{t-1} + \gamma_3 (E - E^*)_{t-1} + \varepsilon_{t+1} \quad (2.10)$$

Model 1b

$$\ln D_{t+1} = \beta_0(t+1) + \sum_{i=1}^{12} \beta_i I_i + \sum_{i=0}^p \delta_i \ln D_{t-i} + \gamma_1 \ln \left(\frac{D}{P}\right)_{t-1} + \gamma_2 \ln \left(\frac{E}{P}\right)_{t-1} + \gamma_3 \ln(|E - E^*|)_{t-1} + \varepsilon_{t+1} \quad (2.11)$$

The class of models 1 is an ARMAX where the exogenous variables are constructed using macroeconomic indicators. $\frac{D_t}{P_t}$, $\frac{E_t}{P_t}$ and $(E - E^*)_t$ represent the dividend-price ratio, the earnings-price ratio and the Okun gap.⁷ I represents the month dummy. The Okun gap is equal to the deviation of actual earnings from the trended mean earnings. In the case of the logarithmic specification, the logarithm

is applied to the absolute deviation. t is the trend dummy. If prices have a trend, then it follows from present value arguments that dividends may have a trend as well. The dividend-price ratio and earnings-price ratio are used as regressors since they are known to forecast returns which are made up of dividends. Intuitively, these variables move because of their ability to forecast returns.⁸ The three variables reflect the possibility that the market has superior information than autoregressive processes. Prices take into account

⁷The Price-dividend and Price-earnings ratio are positively correlated, and it may lead to incorrect variances. However the need here is for forecasting. The Price-earnings ratio exhibits higher variability that may capture very high and very low values of dividend forecasts.

⁸A nice exposure to why the price dividend and price earnings ratio moves is explained in Cochrane (2002).

this market phenomenon. If dividends grow faster, prices will pick it up. Including both $\frac{D_t}{P_t}$ and $\frac{E_t}{P_t}$ may lead to redundancy of one of the variables according to simple t-ratios.⁹ However, variation in the earnings-price ratio is larger than that the dividend-price ratio and hence may forecast extreme events for dividends. The autoregressive lags are inserted only for statistical purposes. The optimal number of lags (p) in the autoregression part is chosen by the Akaike criteria on the differenced form of the regression model (due to the presence of unit roots).¹⁰

Model 2a

$$D_{t+1} = \beta_0(t+1) + \sum_{i=1}^{12} \beta_i I_i + \sum_{i=0}^p \delta_i D_{t-i} + \varepsilon_{t+1} \quad (2.12)$$

Model 2b

$$\ln D_{t+1} = \beta_0(t+1) + \sum_{i=1}^{12} \beta_i I_i + \sum_{i=0}^p \delta_i \ln D_{t-i} + \varepsilon_{t+1} \quad (2.13)$$

Model 2 is a nested form of Model 1, where the corresponding gamma parameters (γ) are equal to zero. If the parameters γ_1, γ_2 and γ_3 are jointly equal to zero, it would mean that market based information has no superior power in predicting dividends.

Model 3a

$$D_{t+1} = \beta_0(t+1) + \sum_{i=1}^{12} \beta_i I_i + \varepsilon_{t+1} \quad (2.14)$$

⁹There is a correlation of 0.71 between the dividend yield and earnings yield. This may lead to statistical rejection of one of the parameter values. However the objective is to predict and high correlation will have no impact on the forecasts since the estimates of the parameters are consistent.

¹⁰In the forecasting literature, the Schwartz Information Criteria (SIC) is the criterion which is mostly used in order to determine the number of lags since it is more penal to overfitting. It should be acknowledged that both the SIC and AIC stipulate the same number of lags in the case of recursive windows.

Model 3b

$$\ln D_{t+1} = \beta_0(t+1) + \sum_{i=1}^{12} \beta_i \ln I_i + \varepsilon_{t+1} \quad (2.15)$$

Models 3 are simple models which state that dividends may be forecast from neither the autoregressive components nor financial ratios.

2.3.2 Stochastic Discount Rate and Dividend Growth

In this section, four models are presented in order to estimate the discount rate defined by (2.9). The denominator (2.9) is made up of two elements, the discount factor and the expected dividend growth. The denominator is probably the most important component of the present value equation. Minor changes in the discount rate and the dividend growth may lead to big changes in the fundamental price, influencing the decision of going long on either bonds or equity. In applied work, the empirical measure used is usually the historical average of realized returns. On the empirical side, the discount rate is computed using moving windows of the realized returns and a cointegration framework proposed by Fama and French (2002). The dividend growth is computed using recursive averages. The computation of the discount and growth rate is explained in the following paragraphs.

The rolling discount rate is the moving average of realized returns over time for a specific period. The estimation window is 30 years for both monthly and annual data. For instance, the monthly average returns over time at a particular date t will be average returns over the past 360 observations (Equation (2.16)).

Model A : Rolling Discount Rate

Model A is the rolling average of realized returns. It is computed by 2.16:

$$E(r_{t+1}) = \frac{1}{360} \sum_{i=t-360+1}^t R_i \quad (2.16)$$

where R_i is the realized real returns on the market and t is the terminal observation where $t > 360$. This definition of discount rate is equal to the average of realized returns since the past 30 years.

Model B: Recursive Discount Rate

The recursive discount rate is the average of realized returns from the beginning of the sample (January 1871) until time t . At time t , the point definition of the expected discount rate is given by (2.17).

$$E(r_{t+1}) = \frac{1}{t} \sum_{i=1}^t R_i, \quad (2.17)$$

where $t \geq 360$. As t increases, the discount factor includes all the data points in the sample starting from the first observation.

Cointegration based Discount Rates

Models (2.16) and (2.17) are historical measures of discount rates. Historical measures are however very noisy, due to the capital gain element in the discount rate.

$$E(r_{t+1}) = \frac{1}{t} \sum_{i=1}^t \frac{D_i}{P_{i-1}} + \frac{1}{t} \sum_{i=1}^t \frac{P_i - P_{i-1}}{P_{i-1}}, \quad (2.18)$$

Equation (2.18) illustrates the average return being decomposed between the dividend yield and the capital gain. The proposition of Fama and French is that if the dividend price ratio (or earnings price ratio) is stationary, then the compound dividend(earnings) growth

approaches the compound rate of capital gain. The intuition is simple. It is a measure of computing expected returns when valuation ratios are fixed. Consider the simple one period return:

$$\begin{aligned} r_{t+1} &= \left(\frac{1 + P_{t+1}/D_{t+1}}{P_t/D_t} \cdot \frac{D_{t+1}}{D_t} \right) \\ &= \left(\left(\frac{D_t}{P_t} + \frac{P_{t+1}/D_{t+1}}{P_t/D_t} \right) \left(1 + \frac{D_{t+1}}{D_t} \right) \right). \end{aligned}$$

r_{t+1} may be approximated as follows:

$$r_{t+1} \approx 1 + \Delta(P/D)_{t+1} + \frac{D_t}{P_t} + \frac{D_{t+1}}{D_t}.$$

This equation explains returns in terms of the dividend yield, and the price increase over current dividends. In the long run, if the price dividend ratio does not change, then $\Delta D = \Delta P$. In other words, the dividend growth is equal to the capital gains element. The model, therefore, requires the assumption that the price-dividend (earnings) ratio is stationary.¹¹ In the long run, the dividend yields revert such that any change in the price-dividend ratio should have a small contribution to mean returns. The equilibrium relationship in this case can be written as:

$$R = D/P + \Delta D$$

In this context, the average stock return or discount rate at time t is the sum of the average dividend yield and average rate of capital gain.¹² The dividend growth model and earnings growth model are written in equations (2.19) and (2.20) respectively.

Model C: Dividend Growth Model

¹¹The price-dividend (earnings) ratio being stationary is similar to price and dividend(earnings) being cointegrated.

¹²This may be easily derived from the definition of returns.

$$E(r_{t+1}) = \frac{1}{t} \sum_{i=1}^t \frac{D_i}{P_{i-1}} + \frac{1}{t} \sum_{i=t}^t \left(\frac{D_i - D_{i-1}}{D_{i-1}} \right). \quad (2.19)$$

Model D: Earnings Growth Model

$$E(r_{t+1}) = \frac{1}{t} \sum_{i=1}^t \frac{D_i}{P_{i-1}} + \frac{1}{t} \sum_{i=1}^t \left(\frac{E_i - E_{i-1}}{E_{i-1}} \right). \quad (2.20)$$

The two models may be linked to the Campbell and Shiller (1988) cointegration framework, where dividend-price ratio and earnings-price ratio vary over time because of the variation in the expected stock returns, expected dividend or earnings growth. Since stock returns and growth rates appear to have constant unconditional means, the dividend-price ratio and earnings-price ratio may be stationary. This is simply because any movement in the dividend-price ratio is explained by the expected returns and dividend growth. In other words, dividend (earnings) and price are cointegrated. The recent literature in financial economics has started analyzing whether the price-dividend ratio is stationary. Some authors such as Lettau and Van Nieuwenburgh (2007) and Campbell and Yogo (2006) believe that the price-dividend ratio may be nonstationary but it is not explosive. This argument makes sense in the presence of bubbles (Diba and Grossman 1988).

However, I define the concept of "fluctuating periods of stationarity" to reconcile the mixed empirical findings. Events such as breaks may cause conventional unit root tests to reject the null of stationarity, when the process is indeed stationary. Tests of stationarity depend on the sample size adopted. To account for the latter, rolling and recursive window tests of unit roots are reported in A.1.9. Interestingly, the rolling tests show periods when the null

hypothesis is rejected. There are also periods when the null hypothesis is not rejected. According to the graphical plots, tests of stationarity for the dividend-price ratio depend on the test sample. Over the full sample, dividend and price, and earnings and price are cointegrated.¹³

The other element of the denominator is dividend growth, which is computed recursively as follows:

$$g_t = E(\Delta d_{t+1}) = \frac{1}{t} \sum_{i=1}^t \ln\left(\frac{D_i}{D_{i-1}}\right). \quad (2.21)$$

It is important to see the implication of each of the four measures of the discount rate on the denominator. The denominator of the present value may be summarized by the following four models. The expected dividend growth from (2.21) is subtracted from (2.16), (2.17), (2.19) and (2.20), to yield the following denominators

A:

$$r_t - g_t = \frac{1}{360} \sum_{i=t-360+1}^t R_i - \frac{1}{t} \sum_{i=1}^t \ln\left(\frac{D_i}{D_{i-1}}\right). \quad (2.22)$$

B:

$$r_t - g_t = \frac{1}{N} \sum_{i=t-N}^t R_i - \frac{1}{t} \sum_{i=1}^t \ln\left(\frac{D_i}{D_{i-1}}\right). \quad (2.23)$$

C:

$$r_t - g_t = \frac{1}{t} \sum_{i=1}^t \frac{D_i}{P_{i-1}} + \frac{1}{t} \sum_{i=t}^t \left(\frac{D_i - D_{i-1}}{D_{i-1}}\right) - \frac{1}{t} \sum_{i=1}^t \ln\left(\frac{D_i}{D_{i-1}}\right) \quad (2.24)$$

D:

¹³The respective t-statistics on the residuals are -4.10 and -4.33.

$$r_t - g_t = \frac{1}{t} \sum_{i=1}^t \frac{D_i}{P_{i-1}} + \frac{1}{t} \sum_{i=t}^t \left(\frac{E_i - E_{i-1}}{E_{i-1}} \right) - \frac{1}{t} \sum_{i=1}^t \ln\left(\frac{D_i}{D_{i-1}} \right). \quad (2.25)$$

Denominator A uses the most recent information on returns while B uses the whole history of information. Denominator C is simply the dividend yield with a term which measures the difference between dividend growth computed from the arithmetic and logarithmic specification. This is the definition used in Bulkeley and Tonks(1989). Denominator D is a model that takes into the speed of growth of dividend and earnings. If the historical average of earnings growth is higher than dividend growth, dividends are discounted at a higher rate than model C. If both earnings and dividends share the same level of growth, Model D is nested within model C.

2.4 Data and Results

2.4.1 Data

Monthly and annual series of real S&P 500 Dividend, Price, Earnings, T-bill rates and Market returns for the period January 1871 to December 2007 are retrieved from Robert Shiller's website. Rolling and recursive window forecasts are generated from January 1901 to December 2007. The initial estimation sample is set from January 1871 to December 1900. The results section is organized as follows. Two sets of results are reported for monthly and annual data. The focus is on monthly data where the strategy performed better.

2.4.2 Monthly Frequency

Table 2.1 illustrates that the strategy is highly profitable over the 107 year sample if investors rebalance their portfolio according to the rule.

Table 2.1:

Annual Rates of Return.

Table 2.1 shows the annual rate of return for the period January 1901 to December 2007. The annual rate for the Buy and Hold strategy is 7.3%. If 100 pounds were invested back in January 2001, the terminal wealth from the Buy and Hold strategy will be £733,213. The combination which yields the highest return is Forecast Model 2 with recursive windows with discount rate A.

Model	A	B	C	D
1a Recursive	9.92	8.93	5.78	6.76
1a Rolling	9.96	8.83	5.78	6.76
2a Recursive	9.96	8.83	5.78	6.76
2a Rolling	9.89	8.96	5.63	6.66
3a Recursive	6.15	6.18	4.70	5.77
3a Rolling	7.21	7.03	5.49	5.71
1b Recursive	9.85	9.15	5.76	6.55
1b Rolling	9.85	9.15	5.76	6.55
2b Recursive	9.84	8.92	5.62	6.55
2b Rolling	9.84	8.93	5.62	6.55
3b Recursive	6.46	5.98	5.07	5.77
3b Rolling	7.58	7.34	5.59	5.85
Buy and Hold	7.3			

Table 2.1 illustrates the annual rate of return if wealth were invested back in January 1901, and allowed to be continuously compounded at the rate of return which the rule postulates ($R_{tr,i} = R_m$ if $P_t < P_t^*$ and $R_{tr,i} = R_f$ if $P_t > P_t^*$). Based on the figures, the best forecast models are models 1 and 2 with discount rate A. Forecast Model 3 does not yield superior profits to the Buy and Hold on average. The functional form of the regression models (linear or logarithmic) does not affect the profitability of the best models.

There is no tractable difference between the recursive and rolling window performance. This may be explained through the fact that the rolling window is large. Based on accumulated returns, the ranking of the best measure of discount rate is A, B, D and C.

The total compounded monthly returns for periods of 24, 36, 48 and 60 months are reported in table A.1 in the appendix. The table shows that the rule works well for shorter horizons as well. The performance of the rule tends to vary for the different forecast models, and definitions of discount rates. The trading strategy beats the Buy and Hold strategy 48%, 58.3 %, 56.25 % and 56.25 % of the time for the 24, 36, 48 and 60 months' horizons respectively. The ranking of the different forecasting models and discount rates is uniform across the different horizons. As the compounding horizon increases, the difference in accumulated returns across horizons tends to increase across the forecasting models and denominators.

In table (A.2) in the appendix, a test of difference in correlated means is reported. The rule significantly outperforms Buy and Hold for longer horizons. For the one period horizon, Buy and Hold is higher than the trading rule return, with the rule beating the market 16 times compared to 20 times in which the opposite happens. As the horizon expands, the rule starts dominating Buy and Hold.

The compounded annualized return rate under the simple Buy and Hold strategy is 6.32 %. The best forecasting model with the best discount rates (Models 1 and 2 with discount factor A) yields a return of approximately 11 %. The rule beats the market 29 times. The best forecast models are models 1 and 2 where they actually beat the Buy and Hold under all discount rates except C. The best performing discount rates are A and B. While the trading rule seems to work in the case of the two best discount rates, it does not recommend switching to the bond market when equity returns

are actually negative. Examples are 1964, 1976 and 2003. If the rules had correctly predicted that the equity market was overpriced during those periods, higher wealth could have been achieved by holding bonds. Two graphical illustrations (figures 2.2 and 2.1) are selectively chosen in order to show how the rule fares under two extremes.

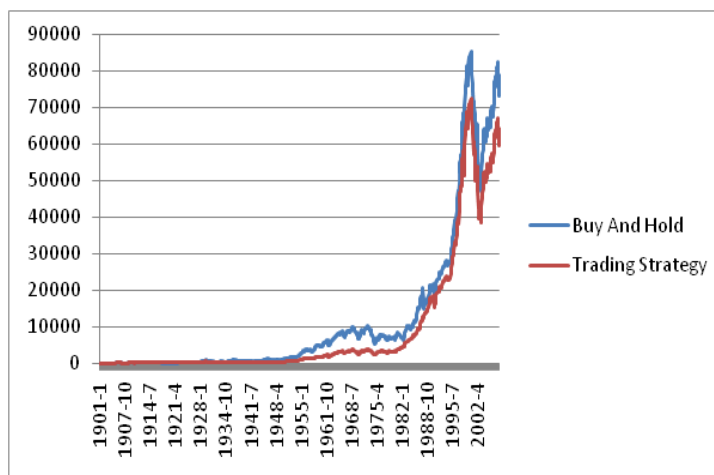


Figure 2.1: **Accumulated Returns for Forecast Model 3a and Discount factor A.** The figure shows the time series plot of accumulated returns under Buy and Hold and the Trading Strategy for Forecast Model 3a and discount factor A.

Figure 2.1 shows one of the worst case scenarios. The strategy picks up the bearish state of market from the 1944-1950. Afterwards, it takes into account the growth of the stock market from the 50's to 70's, where the trading strategy postulates going long on the stock market, most of the time. From the 80's onwards, the strategy postulates going long on equity mostly. However the strategy does not beat Buy and Hold. Figure 2.2, on the other hand, shows that the rule beats Buy and Hold at the terminal date. Throughout the sample, it appears to consistently outperform Buy and Hold.

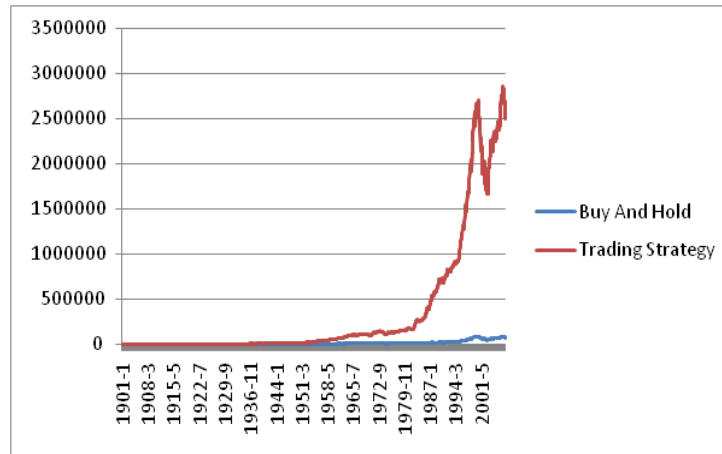


Figure 2.2: **Accumulated Returns Forecast model 1a and Discount factor A.** The figure shows the time series plot of accumulated returns under Buy and Hold and the Trading Strategy for Forecast Model 1a and discount factor A.

Models 1 and 2 are hard to tract through time as they posit many independent switches within the same year. There is no clear trend in a particular holding position for more than 2 or 3 years. On average, switching between bonds and equity may occur 3-5 times in the ARMAX and AR(p) models, as opposed to only once for model 3. Both models advised going long on bond markets from 1917 to 1927. The strategy postulates holding safe assets in the aftermath of the 1916-17 crash. Equity is held during the period 1927-31, and in bonds during the recession (1932-36). A very interesting feature of the model is that it advises agents to go long on bonds as from 1938 itself, before the crash of 1939-42. However, there are periods within this period when the strategy advises going long on equity. Nevertheless, they are few relative to the number of times the rule posits a long position in bonds. Incorrect expectations of the market picking up within this period may be a reason for this evidence. And this will automatically translate in lower expectations of higher prices. The growth of the equity markets during the 50's

and 60's is properly anticipated, where assets are held in equity. The strategy, however, does not pick up the bear market of 1973-74. Instead it advises going long on bonds in the aftermath 1975-76 and 1980-84.

The ARMAX and AR(p) models typically use lagged variables in estimating and forecasting. The window length on the other hand is long. As a new observation of dividend enters the information set, it is averaged (360 times in the rolling window and much more in the recursive window). Unless the new signal of realized dividends is large, expectations of a bear market is not properly assimilated by the forecast models in the initial periods. However as new observations get inside the estimation window, these expectations get picked up by the forecasting model. If a smaller window size was taken for the rolling window (less than 360 months), the new observations of dividends would not be discounted as highly, and the rule would have picked up that the market is overpriced and hence advise going long on bonds. This also explains why the rule correctly identifies market crashes which tend to last for long periods.

In the case of the worst model, the strategy postulates a long position in the equity market until 1914. Bonds are again held during the great depression 1932-36. However, it fails to identify periods when the stock market was down in the period 1939-42. Interestingly, it takes advantage of the growth of the equity market during the period 1949-74. It fails to identify both market crashes of 1973-74 and 1981-83. Given this finding, Models 1 and 2 are better in terms of strategy because they pick up crashes and also they postulate more active trading.

Tables A.7 and A.8 in the appendix, illustrate respectively the number of periods wealth is held in equity and the number of times switching takes place. It may be summarized that model 3a and 3b

(recursive windows) postulate nearly the same number of months to go long on equity. However, models 1 and 2 have more switches. In other words, they can pick up smaller trends in the market. A simple regression of number of switches on accumulated wealth showed a positive relationship between both.

Generally, with regards to the discount rate based on the historical averages (rolling and recursive means), they tend to advise long positions in equity more often. When the cointegration based discount rates are used, best performing models 1 and 2 do not switch as often in a particular year. They tend to exhibit periods of dependence, i.e. If the strategy advised going long in the previous period, it is most likely that it is going to advise going long in the next period as well. The discount rate in this case is very small, such that it increases the perceived present value, implying that the market is more underpriced than it may really be.

An intuitive idea emanates from the cointegration based discount rate. If the discount rate (from any of the models) was a proper reflection of agents' true discount rate, the position of holding bonds or equity will vary according to the forecast error for each dividend forecasting model. Among the different models considered, the AR-MAX and AR(p) model have serially uncorrelated errors. Serially uncorrelated errors imply that there are random fluctuations in the equity or bonds holding positions. Given the results, it can easily be inferred that the discount rate from the cointegrating techniques are high enough to offset the forecast error. In other words, although forecasts from models 1 and 2 are more accurate (and hence more likely to under or over forecast realized dividends in the margin), the discount rate is sufficiently low such that the accuracy of the forecast against the true data generating process, does not matter.

Reliability of the Rule

Risk and transaction costs are considered in this section. In terms of risk, there are two types of risk involved. There is the risk of being in the bond market when the stock market takes off, which consists of risk from ‘missed’ opportunities. The second type of risk is the risk of being in the equity market when the latter actually declines. There is no bond market risk as maturity is being matched to one month. Two simple models are used to test for this nature of risk. In the first case, the Sharpe ratio is used. It takes into account the global riskiness of both strategies. However it is biased towards the rule since the amount of time the rule is in the bond market is not considered. When wealth is held in the bond market, it benefits from lower volatility (risk). To that end, the Sweeney statistic is reported to take into account the proportion of time the asset is held in the two markets.

Sharpe Ratio The Sharpe ratio is used to test for the riskiness of the trading strategy. (2.26) and (2.27) show the computation of the Sharpe ratio for the strategy and Buy and Hold returns.

Trading Rule:

$$SR_{tr}(k) = \frac{R^{tr}(k) - R^f(k)}{\sigma(k)}. \quad (2.26)$$

Buy and Hold:

$$SR_{bh}(k) = \frac{R_{bh}(k) - R^f(k)}{\sigma(k)}. \quad (2.27)$$

where R_{tr} relates to the returns under the trading rule over the period, R_{bh} refers to the returns under the Buy and Hold Strategy and k is the horizon the rule is being put to use.

A test of mean differences was performed for the Sharpe ratio

where correlated means are accounted for. In that case, the Z-score is defined as:

$$Z_{l,d}(k) = \frac{\overline{SR_{tr}(k)} - \overline{SR_{bh}(k)}}{S^2_{\overline{SR_{tr}(k)}} + S^2_{\overline{SR_{bh}(k)}} - 2rS_{\overline{SR_{tr}(k)}}S_{\overline{SR_{bh}(k)}}}.$$

Inference was performed from the student t-distribution.

Table 2.2:

Difference in Means for Sharpe Ratio.

The table shows a summary of the difference in means test for the different horizons. The left column illustrates the different hypotheses being tested. For instance, $SR_{tr} > SR_{bh}$ shows the number of times the Sharpe Ratio for the trading strategy is better than the Sharpe Ratio for Buy and Hold. The total number of models (forecast coupled with discount rates) is 48.

	12 Months	24 Months	36 Months	48 Months	60 Months
$SR_{tr} > SR_{bh}$	20	20	19	19	18
$SR_{tr} = SR_{bh}$	9	8	14	9	3
$SR_{tr} < SR_{bh}$	19	20	15	20	27

Tests of mean differences on the Sharpe ratio show that the rule is quite weak in beating the market return. The individual Buy and Hold strategy tends to work marginally better in the first year. In the first year, there is evidence that the rule may yield returns as high for Buy and Hold. However, for the higher horizons, the Z ratios are less likely to fall in the region of indecision. There is no evidence of the Buy and Hold returns exceeding those of the rule the best models. As expected, the conventional cointegrating discount rate models and the constant mean dividend forecast model does not perform well.

Sweeney X statistic Over the holding period, riskiness may emanate from the variations in both the equity and bond market. Hence, we report the Sweeney statistic (1986). The test is computed as follows:

$$X = R_{tr} - (1 - f)R_{bh},$$

$$\sigma_x = \sigma[f(1 - f)/N]^{\frac{1}{2}}.$$

where $(1 - f)$ is the proportion of months in which the investor's wealth is placed in the equity market, N is the number months in the sample and σ is the standard error of the monthly returns under the Buy and Hold strategy. The results are reported in table 2.3.

Table 2.3:

Sweeney Statistic for Trading Strategy.

The table shows Sweeney's statistic (X/σ_x) over the whole sample period for the different models. Inference may be made from the normal distribution.

	A	B	C	D
1a Rec	7.67	6.42	1.76	1.72
1a Rol	7.73	6.29	1.76	1.72
2a Rec	7.73	6.29	1.76	1.72
2a Rol	7.64	6.46	1.54	1.28
3a Rec	2.53	2.58	0.33	-0.38
3a Rol	4.62	4.35	1.53	-0.61
1b Rec	7.58	6.71	1.74	0.85
1b Rol	7.58	6.71	1.74	0.85
2b Rec	7.55	6.41	1.52	0.86
2b Rol	7.55	6.43	1.52	0.85
3b Rec	3.32	2.63	1.01	-0.38
3b Rol	5.15	4.79	1.72	-0.35

The Sweeney statistic demonstrates that the strategy works better conditional on the discount rate adopted. Again, it is found that

discount rates C and D are not favourable to any of the forecasting models. Models 1 and 2 appear somewhat short of the rejection levels in denominator C. However better results are derived for discount rates A and B. Uniformly, the rule works well with ARMAX and AR(p), irrespective of the functional form of the regression model and the type of moving window. The window type is irrelevant as the length of both windows is large enough so that estimated parameters from rolling and recursive windows tend to be similar.

Transaction costs In this section, we introduce transaction costs, which is a substantial issue given the frequency of trades (switches).

Table 2.4:

Annual Rates of Return with transaction costs of 0.5% per switch.

Table 2.4 shows the annual rate of return for the period January 1901 to December 2007 when a 0.5 % transaction cost is included per switch. The post transaction costs return is lower. In this case, only discount rate A yields marginally higher returns than the Buy and Hold strategy.

Model	A	B	C	D
1a Recursive	8.12	7.11	4.96	6.33
1a Rolling	8.14	7.11	4.96	6.33
2a Recursive	8.14	7.11	4.96	6.33
2a Rolling	8.10	7.15	4.92	6.33
3a Recursive	5.63	5.36	4.26	5.45
3a Rolling	6.35	6.04	4.49	5.45
1b Recursive	8.04	8.21	5.08	6.29
1b Rolling	8.04	8.21	5.08	6.29
2b Recursive	8.04	7.42	5.06	6.29
2b Rolling	8.04	7.42	5.06	6.29
3b Recursive	5.13	5.12	4.83	5.38
3b Rolling	6.45	6.67	4.77	5.62
Buy and Hold	7.3			

Transaction costs depend on the amount traded and other brokerage services. In this study, three worst case scenario transaction

costs were considered namely 1%, 2.5 %, 5 %. The comparison is made by netting the trading rule returns with the transaction costs when a switch is made and computing the Sweeney Z statistic afterwards. When transaction costs are accounted for, the number of models beating Buy and Hold falls drastically. In the worst case scenario, when a monthly transaction cost of 5 % is implicitly assumed, only eight models beat the market return, mostly from denominators A and B. The best forecast models (models 1 and 2) yield high returns. Naturally, the cointegrated discount rate does not yield a favourable outcome with any of the transaction cost rates. The detailed results are shown in tables A.4 to A.6 in the appendix.

Sampling with replacement A small sample simulation experiment is also performed to confirm the robustness of the results. It involves picking out random dates from the period 1901 to an end date which is conditional on the horizon of interest. The selection of the random dates is derived from a uniform probability distribution. A vector of dates is generated using a random number generator where an equally sized vector containing elements between zero and one is randomly chosen from the uniform probability distribution. This vector is then multiplied with $n-k$ where n is the end date of the sample and k is the length of horizon (for example $k = 12, 24, \dots, 60$). This procedure is repeated 160 times for convenience. Subtraction of k ensures that returns under the passive Buy and Hold can be calculated for horizon k , if the draw is near the end of the sample.

The results are reported in table A.12 in the appendix. Models 1 and 2 are the best performers, when coupled with denominators A and B. In the one year horizon, the rule beats the passive Buy and Hold only 41 % of the times. However it is worth noting that there are many instances when the rule is equal to the Buy and Hold

return. As the horizon increases, the number of times the rule beats the market return tends to increase as well. The rule is better for horizons of 3, 4 and 5 years.

After considering various forecast models with different discount rates in the context of accumulated returns, risk and transaction costs, it may be easily seen that the forecast models need to be combined with the appropriate discount rate in order to be successful. In particular, either the ARMAX or AR model may be used with rolling and recursive definitions of discount rate. The measure of the discount rate appears to be the most fundamental issue in computing the present value. Simple rolling and recursive averages of past returns proxy tend to identify perfectly when the market is underpriced or overpriced.

2.4.3 Annual Frequency

When the trading strategy is adopted on an annual frequency, the trading rule does not outperform the Buy and Hold strategy at the terminal date. The cumulated returns are shown in table 2.5.

Application of the rule on an annual basis leads to a much lower terminal wealth than when applied on a monthly basis. The monthly sample allows more compounding (switches). The findings of the rule show that the returns from the Buy and Hold Strategy exceed that of the best trading strategy (Model 1b recursive) by almost two times. Only models 1 and 2 were reported given their success for the monthly data. The strategy performs better for most of the sample involved. However, from 1985 onwards, it advises to go long on bonds. Hence it misses on the bullish markets in 85-86 and 95-99.

Based on accumulated return, the discount rate ranks as A,B,C and D. Compared to the monthly case, there are significant dif-

Table 2.5:

Accumulated Wealth for Annual Data.

The table reports the compounded return from January 1901 to December 2009 for forecast models 1 and 2. For the annual data, the end sample was increased to 2009. The average returns for each discount rate is also reported. The wealth from Buy and Hold is £523,940.

Forecasting Model	A	B	C	D
Model 1(a) Recursive	187,250	234,770	47,020	6,230
Model 1(a) Rolling	195,650	205,220	58,570	6,230
Model 1(b) Recursive	185,710	290,840	87,260	6,230
Model 1(b) Rolling	195,650	216,200	98,230	6,230
Model 2(a) Recursive	298,900	84,710	28,300	4,530
Model 2(a) Rolling	222,170	60,900	28,300	4,530
Model 2(b) Recursive	236,790	109.44	49,820	4,530
Model 2(b) Rolling	287,060	60,900	37,740	4,530
Average Return	226,150	157,870	57,110	5,380

ferences between the ARMA and ARMAX models in terms of the forecast errors. Model 1, which includes market information from the price-dividend, price-earnings and Okun's gap seems to work well when the discount factor is computed on a rolling window basis (B) and also for discount rates C and D. The autoregressive models work better only when the window size is large.

The results are compared with the filter rule of Bulkley and Tonks (1991). The filter rule involves comparing the price to the average price (a proxy for the long run rational expectations price) and choosing the holding position based on this. This measure does not inherit from forecasting errors. For simplicity, we reproduce the holding position of Bulkley and Tonks (1991). The trading strategy is not replicated for years after their end of sample size which is 1985, because their computation of the discount rate is done recursively backwards. Selection of the terminal year 2009 may not lead to the same discounted value. Figure 2.3 shows the accumulated

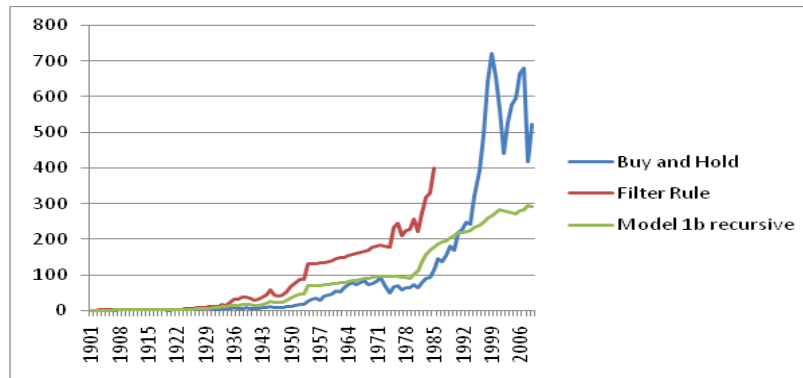


Figure 2.3: **Accumulated Returns under Buy and Hold, Trading Strategy and Filter Rule.** The figure shows the plot of compounded returns for Trading Strategy (green line), Buy and Hold (blue line) and Filter Rule (red line) over time.

returns from the different strategies.

Figure 2.3 shows that both the filter rule and the trading strategy beat Buy and Hold in the initial stages of the sample in 1922. However the filter rule tends to pick up early in the sample around 1903-04 and 1909-13 respectively. The graph shows that the difference starts to become important as from 1929 itself. The trading strategy tends to be higher than the equity index until the year 1986. Before that period, the filter strategy tends to work better than the trading strategy and Buy and Hold. In 1985, the filter strategy has a terminal wealth which is nearly 3 and 2 times more than Buy and Hold and the trading strategy respectively. The strategy after the 1985 period states a position in bonds, and hence is insulated from the 1999 and 2007 crash. However again, it fails to take into account the growth of equity in the period 85-95.

An in-depth comparison between the filter rule and the trading strategy shows that the latter is more conservative and suggests that bonds should be held. The difference in the holding positions between the two strategies occurs when the filter rule stipulates going

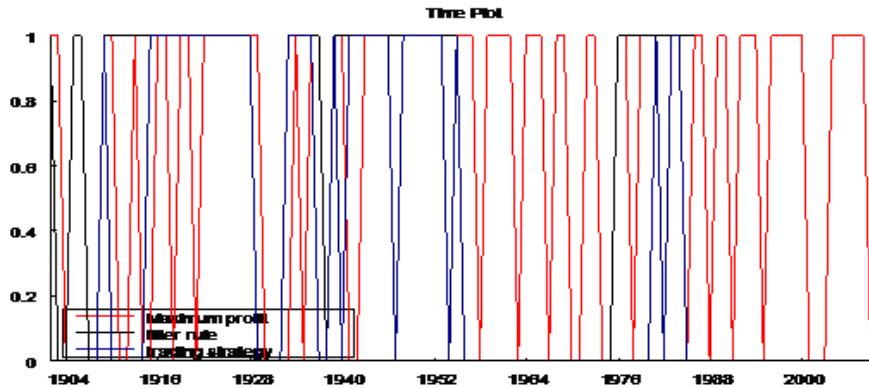


Figure 2.4: **Holding Positions for Filter Rule, Trading Model and Optimal Profit.** The red line shows the holding positions which should be adopted for achieving the optimal profit. The trading strategy and the filter rule are represented by the blue and black line respectively.

long on equity. This is observed in 1904-05, 1909-13, 1936, 1946, 1953 and 1975-79 and 1984-85. The filter rule correctly identified periods when the equity market boomed, except in 1946 and 1953. Thus the forecast models are more risky in incorrectly identifying bull markets. However, the other type of risk should also be mentioned; in particular, whether the rule posits a position in equity when the equity market is falling. The forecast model can be seen to perform better in this case.

Figure 2.4 shows the holding positions for the filter and forecasting rules. A position of 0 simply means that the asset is held in bonds, while 1 implies the asset is held in equity. The ‘Maximum Profit’ series shows the holding position which makes the maximum profit looking back in time. It is the maximum return on the risk free rate and the equity market return in each period. Interestingly, until the period 1985, the maximum profit involves switching a lot between bonds and equity over the sample. Until 1985, it suggests that wealth should be held in equity 57 times (compared to 56 and 36

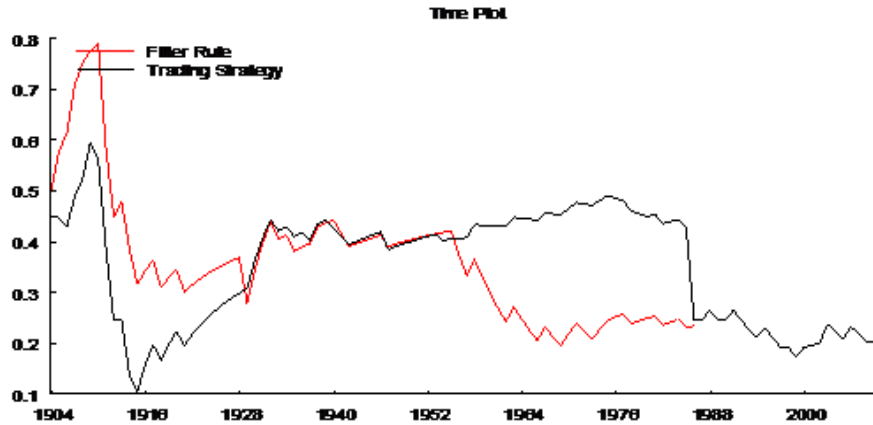


Figure 2.5: **Correlation of Holding Position for Filter Rule and Forecast Model with Ideal Profit.** The figure shows the movement in the correlation between each strategy and the optimal holding position. The correlation for the filter rule and the trading strategy are illustrated by the red and black line respectively.

respectively for the filter rule and the forecasting model). However, 2.4 does not necessarily show the filter tallies with the maximum profit in terms of providing the correct position held.

Figure 2.5 shows that as the horizon increases the holding position between the forecasting rule and the maximum profit overtakes the filter rule. The latter starts off better in the initial periods. However, after 1918 the strategy picks up the bullish states and starts positing the correct position. The forecasting strategy finally starts overtaking the filter rule in 1930. It makes the most of the equity market equity growth from 1940-54. Due to the incorrect position during the 60's, the correlation falls drastically in 1985.

The cumulated returns over the different years for the Buy and Hold and the Trading Rule are reported in table A.13. In this case, the decision of going long on the equity or bond market is made at a specific time and the asset is held for the length of the horizon.

For instance, if the rule posits a long position on bonds at time t , wealth is held in bonds until $t + k$, where k is the horizon length. The results for the holding periods show that returns of the rule do not tend to exceed the Buy and Hold Strategy. As the horizon expands, the cumulated returns from the Buy and Hold increases and falls again. It is important to note that reversion to the present value may take different years in each potential reversion.

There are reasons as to why the trading strategy may not be successful for the annual frequency. The simplicity of the model is that it assumes reversion to the discounted present value. However, in financial markets, actual prices may never revert back to a previously defined ‘equilibrium’ as expectations change over time. At every point in time, a new present value is formed which indirectly includes speculators’ own discount rates. Based on the annual results, the present value states that the stock market is overpriced most of the time, and hence the asset should be held in bonds.

2.5 Conclusion

This chapter points out interesting results not previously evidenced in the literature. First of all, there is a possibility of present value reversion happening at higher frequencies than annual data. Higher cumulated returns may be earned by applying a rule which arbitrages away any opportunity offered by the mispricing of equity returns. The switching frequency is very important. Application of the rule to monthly data leads to higher profits than when adopted on annual data. The strategy appears to yield higher returns for holding horizons of more than one year. However, it tends to be sensitive to the forecasting model and the discount rate adopted. For monthly data, both AR(p) and ARMAX models work well with

the historic averages of returns (discount rate). The model works well both in Bullish and Bearish states.

For the annual data, the rule performs moderately in real-time. It shows the market as overpriced more often and hence postulates going long on bonds. However, it misses on bull markets after 1985. A very important point worth highlighting is that the rule is insulated against bearish positions in the market. However, it misses out on periods where equity returns are high. There are two types of risk involved in the strategy. There is the risk of being in bonds, when the stock market picks up. There is also the possibility of being in the stock market when it goes down. Since the conventional Sharpe ratio does not differentiate between both types of risks, the Sweeney statistic was used to test whether the strategy is profitable. The strategy passes this first pass test. However, the returns are much lower once transaction costs are taken into account. The best model beats the Buy and Hold strategy by approximately only 2 %. Another test considered in the paper is the replacement sampling test. The test randomly draws an initial date from the sample and applies the strategy to that sample. Various holding positions are considered. The replacement sampling procedure shows that the strategy marginally outperforms the Buy and Hold.

Perhaps the most important idea that this chapter conveys is the computation of discount rate. In a recent American Financial Association presidential address, Cochrane (2011) mentions the importance of discount rates. This may be clearly illustrated in our trading strategy. Forecast models are important, but if they are not used with the proper definition of discount rate, the present value strategy appears futile. The two categories of discount rates considered were the historical average and also a discount rate based on fundamentals. The latter tends to be lower. The economic intuition for this approach is simple. The discount rate based on fundamen-

tals is a long run measure and may not reflect the discount rate measured at monthly frequencies.

The current research may be improved by looking at individual securities, or different stock markets. The work of Bulkley and Tonks(1989 and 1992) considered the UK and US markets respectively. Bulkley and Taylor(1996) applied to the model to individual securities. The current research uses monthly frequency data, real-time forecasting and uses a replacement sampling test, which has not yet been explored in the literature.

Chapter 3

Trading Rule Based on Latent Variable Approach

3.1 Introduction

In the previous chapter, the mean reverting strategy was implemented using econometric forecasts of discount rates, dividends and dividend growth. In this chapter, a structural model is put forward to derive these series. The present value is formulated in a state space model. The optimization of the state space model yields a vector of present value parameters which is used to compute time series of expected returns and expected dividend growth. The present value of the market index is constructed using the derived series. The present value is then compared to the actual price to decide whether to go long on equity or bonds.

The real-time nature of the model is taken into account by the Kalman filter whereby each observation is forecast recursively. The model is based on Koijen and Van Binsbergen (2010) who use state space modeling to derive the expected returns and growth in a present value framework, where the Kalman Filter is used to derive the log likelihood function. The intuition behind this methodology

is that expected returns and expected dividend growth are unobservable to the econometrician. However we do observe the realized values. In each recursion, the agent seeks to minimize the forecast error between realized and expected values.

The expected returns and expected growth series are filtered from realized observations based on the Kalman procedure, where expectations are updated as a new observation of the realized value is known. The law of motion for the price-dividend is derived from the Campbell and Shiller(1988) approximation, assuming that expected returns and the dividend growth rate follow an autoregressive process. The state space model is derived from the present value relationship between price-dividend ratio, expected returns and expected dividend growth. The Kalman Filter is applied to the model parameters which are optimized using the conditional Maximum Likelihood procedure.

The present value approximation states that the price-dividend ratio is the link between expected returns and growth. The structural decomposition of expected returns and growth makes use of this approximation. The build of the state space model involves the price-dividend ratio and realized dividend growth as being the measurement equation with the expected returns and expected dividend growth processes forming the state equation. The intuition is that the unobserved variables (expected returns and growth) are related to measured variables. In terms of econometric estimation advantages, the Kalman Filter provides estimates that are robust to structural breaks and does not require the estimation of a large number of parameters (Rytchkov 2007).

After deriving the expected returns and expected dividend growth series, the present value of the index may be easily computed. The present value is compared to the actual price to decide whether to

go long on equity or bonds. If the present value is higher than the actual price, implying the market is underpriced, the proper strategy is to hold equity. On the other hand if the market price is high relative to the present value, holding equity would lead to a capital loss and the proper strategy is therefore to shift the asset from equity to bonds.

3.2 Literature Review

This section reviews studies related to the trading strategy, present value and applications of state space models.

3.2.1 Trading Rule

The theoretical underpinning of the rule involves the comparison of stock price to the present value in order to define the holding position in either bonds or equity. Theoretically, the present value is determined by aggregated expectations of agents and the type of process they used to model the data generating process of dividends (Timmermann 1993). Any difference between the present value and the actual price offers possibilities of arbitrage. In the next period(s), this difference may be arbitrated away as agents revise their expectations. The revision of expectations may occur two ways. Agents may either change the parameters in their forecasting model or change the forecasting model itself.

A profitable opportunity may arise during the adjustment of the market price towards the fundamental value. The rule defines going long on the bonds if the market is overpriced, and equity otherwise. A brief review of three related studies is given in chapter two. Bulkeley and Tonks (1989, 1991) apply the trading rule for the UK and

the S&P 500 equity index. The main premise was to compare the ex post rational price with the present value. The comparison can then be used to define whether the market is overpriced. In both cases, returns from the strategy exceeded the returns from Buy and Hold. The trading rule returns are marginally better than Buy and Hold when risks and transaction costs are taken into account.

3.2.2 Present Value

The present value model is based on a general equilibrium economy where any riskless return may be arbitrated away. The basic premise of the present value is that the theoretical price is derived by the infinitely discounted payoffs from the asset. The present value is the same as (2.4) in chapter two:

$$P_t^* = E_t \sum_{i=1}^{\infty} [r_{t+i} D_{t+i}], \quad (3.1)$$

where E_t is the expectations operator at time t , r_{t+j} is the return from time t to $t + j$ and D_{t+j} relates to the dividend at time $t + j$. When dividends grow at the rate of Δd_{t+1} , the new present value formulae is equal to (3.2).

$$P_t^* = \frac{1}{E_t[r_{t+1} - \Delta d_{t+1}]} E_t[D_{t+1}]. \quad (3.2)$$

(3.2) is used to compute the present value in this chapter. For interesting applications of the present value, see Cuthbertson (2002), Kanas (2005), Rangvid (2006), Shiller and Beltratti (1993), Allen (2004), Caporale and Gil-Alana (2004), Strauss (2001), Bohl and Siklos (2004), Mills (1993).

3.2.3 Time Varying Expected Returns and Dividend Growth

(3.2) may be rewritten such that the present value is an approximation which relates the price-dividend ratio to the expected returns and expected dividend growth (Campbell and Shiller 1988). This shall be explained more thoroughly in the methodology section.

The objective of applying the state space model is to derive expected returns and expected dividend growth which are both latent but are linked through the price-dividend ratio approximation. If prices and dividends are cointegrated¹, then all the variation in the price-dividend ratio must come from the variation of expected returns and dividend growth. It is generally popularized in the literature that dividend growth is unpredictable (Cochrane 1992), Lettau and Ludvigson (2001), Cochrane (2008), Lettau and Van Nieuwerburgh (2008)). Therefore, all time variation in the price-dividend ratio comes from the expected returns. To derive the series of expected returns, the state space model is used, such that the state equations describe potential linear variations in the expected returns and dividend growth.

Several models have been used to proxy expected returns and growth such as the simple trend, predictive Ordinary Least Squares, Bayesian models and State Space models. Since both expected returns and expected dividend growth rate are unobservable, the state space model can be used to provide efficient estimates of these two variables given observed data. The Kalman filter has been used in the literature to uncover expected returns. Conrad and Kaul (1988) apply the Kalman Filter to extract expected returns from the history of realized returns. The objective was to attempt to characterize the random nature of expected returns and test whether the latter was constant. Brandt and Kang (2004) investigated the relationship be-

¹It implies stationarity of the price dividend ratio.

tween expected returns and volatility. They model conditional mean and volatility of returns as unobservable variables which follow a latent VAR model and filter them from observed returns. In the same line of thinking, Cochrane (2008) shows that the VAR model can be represented in state space form. Pastor and Stambaugh (2009) use the Kalman Filter to analyze the correlation between predictors and expected return in the forecast of returns. Kojien and Van Binsbergen (2010) use the state space model to model dividends.

3.3 Methodology

3.3.1 Present Value Model

In this section, the present value relationship between the price-dividend ratio, expected returns and expected dividend growth is derived. The series is developed from a theoretical assumption that both expected returns and dividend growth rate follow an autoregressive process of order one².

The rate of return is defined as

$$r_{t+1} = \log\left(\frac{P_{t+1} + D_{t+1}}{P_t}\right). \quad (3.3)$$

The Price-Dividend ratio is defined as

$$PD_t = \frac{P_t}{D_t}. \quad (3.4)$$

The Dividend Growth rate is defined as

$$\Delta d_{t+1} = \log\left(\frac{D_{t+1}}{D_t}\right). \quad (3.5)$$

²The AR(1) has interesting properties. It adequately captures time series properties of expected returns without needing to compute a large number of parameters in the state space model.

One of the important assumptions is the type process of expected returns and dividend growth. The intuitive idea concerning the functional form of the process is that it should be able to illustrate the dynamics of the variables. However, the endeavour of finding a model close to the true data generating process is hectic and involves a lot of data mining. The mean adjusted conditional expected returns and dividend growth rate are modelled as an autoregressive process as in (3.6) and (3.7) respectively:

$$\mu_{t+1} - \delta_0 = \delta_1(\mu_t - \delta_0) + \varepsilon_{t+1}^\mu, \quad (3.6)$$

$$g_{t+1} - \gamma_0 = \gamma_1(g_t - \gamma_0) + \varepsilon_{t+1}^g, \quad (3.7)$$

where $\mu_t = E_t(r_{t+1})$ and $g_t = E_t(\Delta d_{t+1})$. (3.6) and (3.7) shows an

autoregressive process of order one in the mean deviation of the expected returns and expected dividend growth rate. δ_0 and γ_0 represent the unconditional mean of the expected returns and dividend growth respectively. δ_1 and γ_1 represent the autoregressive parameters. ε_{t+1}^μ and ε_{t+1}^g are shocks to the expected returns and the dividend growth rate processes. $\varepsilon_{t+1}^\mu \sim N(0, \sigma_\mu^2)$ and $\varepsilon_{t+1}^g \sim N(0, \sigma_g^2)$. However, no restrictions between the covariance of ε_{t+1}^μ and ε_{t+1}^g is assumed because market shocks will affect both expected returns and expected dividend growth which may have feedback effects on each other.

The realized dividend growth rate is defined as the expected dividend growth rate and the unobserved shock ε_{t+1}^d , where by :

$$\Delta d_{t+1} = g_t + \varepsilon_{t+1}^d. \quad (3.8)$$

ε_{t+1}^d and g_t are assumed to be orthogonal to each other $E(\varepsilon_{t+1}^d, g_t) = 0$.

The Campbell and Shiller (1988) log linearized return approximation (derived in A.2.1) may be written as :

$$pd_{t+1} = \kappa + \rho pd_{t+1} + \Delta d_{t+1} - r_{t+1}, \quad (3.9)$$

where $pd_t = \log(PD_t)$, κ is an arbitrary constant defined as $\log(1 + \exp(\overline{pd})) - \overline{pd}$ and $\rho = \frac{\exp(\overline{pd})}{1 + \exp(\overline{pd})}$. Alternatively, future realized returns may be written as a function of the current and future price-dividend ratio and future dividend growth.

$$r_{t+1} = \kappa + \rho pd_{t+1} + \Delta d_{t+1} - pd_t. \quad (3.10)$$

To study the dynamics of the price-dividend ratio, the process may be written with pd_t being the subject of the formula:

$$pd_t = \kappa + \rho pd_{t+1} + \Delta d_{t+1} - r_{t+1}.$$

By replacing lagged iterated values of pd_{t+1} in the equation, the process may be written as :

$$pd_t = \sum_{i=0}^{\infty} \rho^i \kappa + \rho^\infty pd_\infty + \sum_{i=1}^{\infty} \rho^{i-1} (\Delta d_{t+i} - r_{t+i}).$$

$$pd_t = \frac{\kappa}{1 - \rho} + \rho^\infty pd_\infty + \sum_{i=1}^{\infty} \rho^{i-1} (\Delta d_{t+i} - r_{t+i}).$$

The equation shows that the price-dividend ratio is a constant term and moves according to changes in future realized dividend growth or returns.

3.3.2 State Space Representation

A simple state space model usually has a state equation and a measurement equation. In the present model, there are two state variables (unobserved μ_t and g_t) and two measurement variables, namely

(Δd_t and pd_t). Using the present value approximation, only one state equation is required. The present value model can be shown to include two measurement equations and one state equation. The model parameters are estimated before making the forecasts. The likelihood of the Kalman filter is optimized to derive the parameters of the model. The next paragraphs show the set up of the state space model.

The two transition equations (3.11) and (3.12) show the evolution path of demeaned growth and demeaned expected returns.

$$\tilde{g}_{t+1} = \gamma_1 \tilde{g}_t + \varepsilon_{t+1}^g, \quad (3.11)$$

$$\tilde{\mu}_{t+1} = \delta_1 \tilde{\mu}_t + \varepsilon_{t+1}^\mu, \quad (3.12)$$

where $\tilde{g}_{t+1} = g_{t+1} - \gamma_0$ and $\tilde{\mu}_{t+1} = \mu_{t+1} - \delta_0$. The two measurement equations are given by :

$$\Delta d_{t+1} = \gamma_0 + \tilde{g}_t + \varepsilon_{t+1}^d, \quad (3.13)$$

$$pd_t = A - B_1 \tilde{\mu}_t + B_2 \tilde{g}_t. \quad (3.14)$$

(3.12) can be rearranged with (3.14) such that there are only two measurement equations and only one state equation.

$$\tilde{g}_{t+1} = \gamma_1 \tilde{g}_t + \varepsilon_{t+1}^g. \quad (3.15)$$

$$\Delta d_{t+1} = \gamma_0 + \tilde{g}_t + \varepsilon_{t+1}^d. \quad (3.16)$$

$$pd_{t+1} = (1 - \delta_1)A + B_2(\gamma_1 - \delta_1)\tilde{g}_t + \delta_1 pd_t - B_1 \varepsilon_{t+1}^\mu + B_2 \varepsilon_{t+1}^g. \quad (3.17)$$

(3.15) defines the transition (state) equation. The measurement equation relates the observed to the unobserved variables. In our case this is given by (3.16) and (3.17). A is equal to $\frac{\kappa}{1-\rho} + \frac{\gamma_0 - \delta_0}{1-\rho}$, $B_1 = \frac{1}{1-\rho\delta_1}$, $B_2 = \frac{1}{1-\rho\gamma_1}$. The parameters A , B_1 and B_2 are the present value parameters. The state space equations (3.15), (3.16) and (3.17) are represented in matrix form in the appendix 2. Since all the equations are linear, we can implement the Kalman Filter and obtain the likelihood which is maximized over the following vector of parameters.

$$\Theta = (\gamma_0, \delta_0, \gamma_1, \delta_1, \sigma_g, \sigma_\mu, \sigma_d, \rho_{g\mu}, \rho_{gd}, \rho_{\mu d})$$

3.3.3 Trading Strategy

In this section, we compute the present value from the series of expected returns and expected dividend growth derived in the earlier section. The present value of the equity index is given by (3.2). After replacing the parameters with the new notations for expected values, the present value may be written as (3.18).

$$P_t^* = \frac{1}{\mu_t - g_t} E_t[D_{t+1}], \quad (3.18)$$

where μ_t and g_t are the filtered forecasts of expected returns and dividend growth. In the first chapter $E_t[D_{t+1}]$ was proxied by forecast models. Here, we shall use the dividends until time t and the corresponding derived growth. $E_t[D_{t+1}]$, is computed as:

$$E_t[D_{t+1}] = D_t(1 + g_t).$$

The expected future dividend based on expectations at time t is made up of the previous period dividend compounded with the expected growth rate. The trading rule can be summarized as follows:

Go long on the equity index if:

$$\frac{P_t^*}{P_t} > 1.$$

Go long on the risk free asset if:

$$\frac{P_t^*}{P_t} < 1.$$

3.4 Data and Results

Data on the price-dividend ratio and dividend growth was retrieved from Shiller for the period December 1900 to December 2008. The objective is to derive the series of expected returns and growth. For the application of the strategy, data on market returns and bond returns is collected from the same source. The result from the state space optimization is reported in table 3.1.

It is important to reconcile the optimization procedure with real-time. The optimal parameters from table 3.1 are fixed throughout the sample size. The structural parameters are not time-varying where as the variables which make up the present value are. The conditional forecasts are made on the information set available at that time. However the structural parameters are fixed throughout. The model may be optimized over different sample sizes as a measure of how robust the structural parameters are over time.³ The focus of real-time is that agents update their expectations of returns based on the information. However, the process of updating is believed to be fixed with regards to the present value parameters.

³However this procedure may be computationally intensive.

Table 3.1:

Optimization of Present Value Model.

The table illustrates the solution to the optimization problem explained in section 3.3.2. The optimal values of the parameters of the model and their standard errors are reported in the second and third columns respectively. The standard errors are computed analytically from the Hessian Matrix.

Parameter	Coefficient	Std. error
γ_0	0.0012	0.001
δ_0	0.0405	0.0503
γ_1	0.8056	0.0246
δ_1	0.9988	0.0032
σ_g	0.0066	0.0005
σ_d	0.0023	0.0002
σ_μ	0.0071	0.0003
$\rho_{g\mu}$	0.6006	0.0326
$\rho_{\mu D}$	0.1013	0.0321

From table 3.1, the unconditional mean of expected dividend growth (0.1 %) is lower than that of expected returns (4 %). The standard errors are relatively low in both cases, implying that expectations of the unconditional mean do not vary a lot over time. There is a high persistence in both expected returns and dividend growth. Expected Returns tend exhibit near unit root behavior. Also noticeable is the higher level of volatility of expected returns than dividend growth for the unconditional mean. The autoregressive parameters show more variability in the dividend growth than the expected returns on the other hand. The growth rate is also found to be persistent, which may normally be expected at monthly frequencies. There is also a high positive correlation between the expected returns and dividend growth rate. The positive correlation implies that there will be offsetting effects of the expected returns and expected dividend growth rate on the price-dividend ratio⁴. However shocks to the realized dividend process and the

⁴The positive correlation between both the expected returns and growth ensures that the

expected returns tend to exhibit a low correlation.

3.4.1 Properties of Expected Returns and Dividend Growth

After deriving the optimal parameters, the time series of expected returns and growth variables are retrieved from the present value model. The time series plots of realized and expected returns and dividend growth series are illustrated in figures 3.1 and 3.2. We also report the distributional statistics of the series in table 3.2.

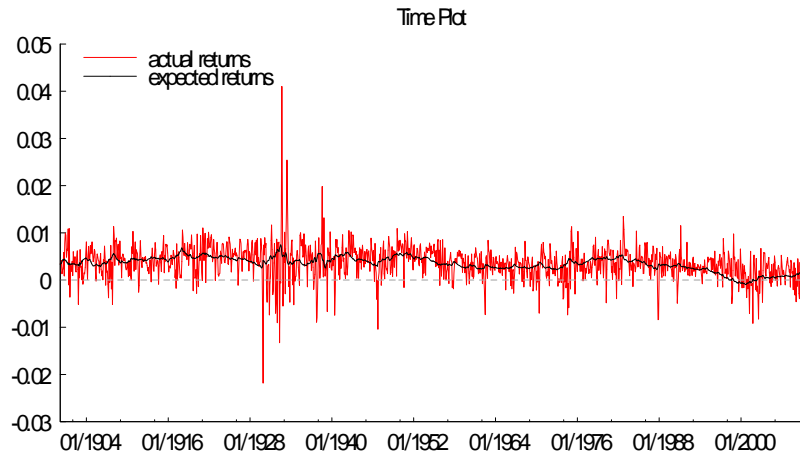


Figure 3.1: **Plot of Expected and Realized Returns.** The figure illustrates the plot of the realized returns (red line) and expected returns (black line) from January 1901 to December 2008.

Figure 3.1 shows the time series of expected and realized returns. Expected returns tend to exhibit serial correlation, which contrasts with the level of correlation in realized returns. However, during stock market crashes expected returns are high and during boom periods, expected returns are low. On the other hand, there seems

net effect between the two latent variables 'moves' the price dividend ratio.

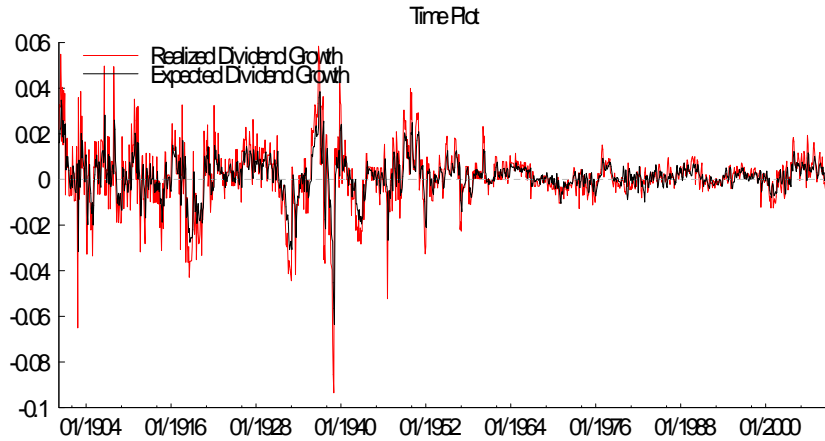


Figure 3.2: **Plot of Expected and Realized Dividend Growth Rate.** The figure illustrates the plot of the realized dividend growth (red line) and expected dividend growth rate (black line) from January 1901 to December 2008.

to be a contemporaneous relationship between realized and expected dividend growth rate. In the model, the expected returns series is freely determined as illustrated from (3.16) and (3.17) (in other words, the expected returns does not enter in the state space model directly). More control is allowed for dividend growth which may be the reason as to why this contemporaneous relationship exists.

Table 3.2:

Summary Statistics for Expected and Realized series.

The mean, standard deviation, and other moments are reported for the realized and expected values of returns and dividend growth.

	\mathbf{R}_t	$\boldsymbol{\mu}_t$	$\boldsymbol{\Delta d}_t$	\mathbf{g}_t
Mean	0.0036	0.0035	9.4×10^{-4}	0.001
Std. Deviation	0.044	0.017	0.013	0.008
Skewness	0.202	-0.812	-0.926	-0.859
Kurtosis	14.41	3.563	9.97	0.11
Jarque-Bera	7105	161	2788	2918

Table 3.2 illustrates the summary statistics for the filtered series. Both the mean and variance of the realized and expected returns tend to be close to each other. However the third moments show that expected returns tends to be highly negatively skewed compared to the positive skewness of the realized returns. Both realized and expected dividend growth tend to be highly negatively skewed. The kurtosis and Jarque-Bera statistic rejects normality of distributions in all the three series. Stationarity tests and correlation across time for the different series are presented in tables A.14 and A.16 in the appendix. The null hypothesis of stationarity is not rejected by all tests in the realized and expected dividend growth series.

In the case of expected returns, the results are, however, mixed. The possibility of expected returns series being nonstationary is reinforced by the nonstationarity tests. The tests do not reject the possibility of a unit root in the series. On the other hand, all the other series (realized returns, expected and realized dividend growth) are found to be stationary⁵. In terms of correlation, both the expected dividend growth and realized returns have diminishing correlation with their past lags.

3.4.2 Trading Rule Returns

In this section, the trading rule results are reported. The performance of the strategy is judged by two simple criteria: end of period wealth and the correct number of times the rule posited the correct position. Figure 3.3 shows the cumulated returns over time for the strategy and Buy and Hold.

⁵Instead of assuming a process as being either $I(0)$ and $I(1)$, a fractional root approach may solve this dichotomy.

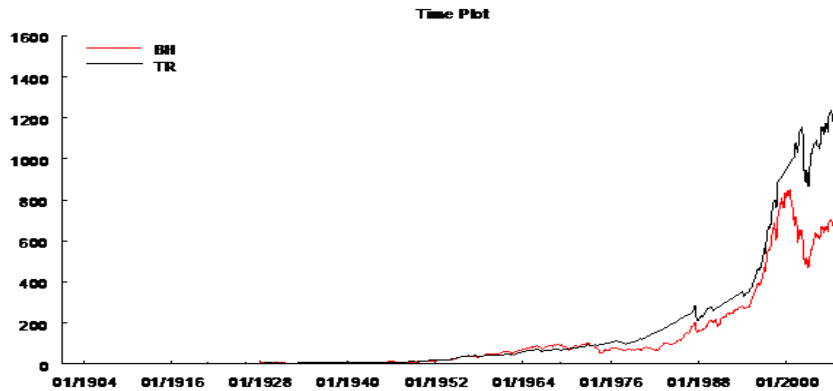


Figure 3.3: **Cumulated Returns from Trading Strategy.** The graphical plot illustrates the cumulative returns over the horizon 1901-2008 from Buy and Hold (BH) and the Rule (TR) if £1 were invested in January 1901.

The rule outperforms Buy and Hold at the terminal date by almost two times. However, over the sample period, Buy and Hold seems to perform as well as the trading strategy. In fact the accumulated return from Buy and Hold tends to be better until 1970. The performance of the trading strategy over the sample is considered in the next paragraphs.

The Buy and Hold performs better than the trading strategy in 1902-03. From 1904 to 1920, Buy and Hold still outperforms the trading strategy. Although the trading strategy correctly postulates to go long on bonds for the period 1914-18, the strategy does not overcome Buy and Hold. The trading strategy overcomes the Buy and Hold during the Great depression 1931-1935 and in 1937-43. However, the rule does not make the most of the booming equity market for the period 1953-73. The rule correctly identifies the overpricing in 73-74 however. From then on, the trading strategy exceeds the Buy and Hold. However, it should be noted that the strategy afterwards postulates going long bonds when it could have made the most of a higher return on the equity market.

The trading strategy postulates going long on equity only 696 times (54 % of the sample size). Portfolio rebalancing happens only 246 times. Compared to the econometric forecasts from chapter one, the state space model also advises less switches than the best models for the dividend forecast. The trading strategy makes the correct decision only 680 times. The trading strategy is biased towards identifying bullish markets and is prone to wrongly identifying the correct state of the market when the latter is bearish. In other words, it is biased towards market overpricing. This may be backed by the finding that out of the correct decisions made, it postulates going long on bonds 405 times compared to only 275 times on equity. Similar to the results in chapter one, the portfolio position exhibits high serial correlation over time. Both strategies have relatively the same standard deviation. The model illustrates that the market is highly overpriced especially during once off events. Figure A.14 in the appendix shows the probability distribution of $\frac{P_t^*}{P_t}$. The figure shows that the distribution is right skewed and has a mean slightly higher than the efficient markets⁶, implying that most of the time the market is overpriced.

3.4.3 Tests on the Trading Strategy

The return and the standard error for 12, 24, 36, 48 and 60 months' horizons are reported in table 3.3. It simply shows the cumulated returns if the strategy was adopted over these horizons.

Contrary to the graphical depiction, the Buy and Hold strategy has marginally a higher return than the trading strategy. As the horizon increases, the return margin between the rule and the equity index tends to increase as well. The standard error tends to remain more or less constant over the k periods. The increasing margin

⁶If markets are efficient, $\frac{P_t^*}{P_t}$ should be close to one.

Table 3.3:

Cumulated Returns over Horizons.

The table illustrates the cumulated returns for 12, 24, 36, 48 and 60 months' horizons for the strategy and Buy and Hold. The standard errors are also reported as measures of the volatility of returns under the trading strategy and Buy and Hold. The cumulated returns for k periods from the Buy and Hold and

Trading Rule are computed as follows $R^{BH}(k) = \frac{1}{T-k} \sum_{h=1}^{T-k} \prod_{i=h-k}^K (1 + R_{m,i})$

and $R^{TR}(k) = \frac{1}{T-k} \sum_{h=1}^{T-k} \prod_{i=h-k}^K (1 + R_{tr,i})$ respectively.

Period	R^{BH}	R^{TR}	S.E ^{BH}	S.E ^{TR}
1 year	0.08	0.07	0.0019	0.0011
2 year	0.17	0.13	0.0035	0.0028
3 year	0.26	0.25	0.0036	0.0028
4 year	0.37	0.35	0.0036	0.0029
5 year	0.48	0.45	0.0036	0.0029

shows that the rule would be better suited for longer horizons since for shorter horizons, Buy and Hold is better after accounting for the variation in returns.

If the volatility measures are taken into account, the trading strategy is better. For one unit of the standard deviation, the trading strategy exceeds Buy and Hold at all horizons, except for the two year horizon. The better performance of the trading strategy is due to the presence of the bearish regimes, which commands a lower standard error.

3.4.4 Robustness of the Rule

In this section, we check the robustness of the earlier results. The earlier results showed that the trading strategy works better over the whole sample size and is marginally better for the different periods.

In this section, the risk of the trading strategy is considered. Three tests are considered namely a simple test of paired correlation, the Sweeney’s X statistic, and a sampling method.

Test of Paired Correlation

We also test whether Buy and Hold is outperformed by the strategy by performing a test of pairwise correlated means on the returns.

$$t(k) = \frac{\overline{R_{BH}}(k) - \overline{R_{TR}}(k)}{S_{R_{BH}}^2 + S_{R_{TR}}^2 - 2rS_{R_{BH}}S_{R_{TR}}},$$

where $t(k)$ refers to the t-statistic for a horizon of k months. $\overline{R_{BH}}$ refers to the mean return on the market (Buy and Hold Strategy) and $\overline{R_{TR}}$ refers to the mean return under the trading rule. $S_{R_{BH}}$ refers to the standard deviation on the market return and $S_{R_{TR}}$ is the market return under the rule with r being the correlation coefficient. The results are reported in table 3.4.

Table 3.4:

Test of Correlated Means.

The right hand side column illustrates the holding period (k). The correlation between the two return series is also reported. $\overline{R_{BH}}(k) - \overline{R_{TR}}(k)$ defines the mean difference. The denominator in the test is given by the standard. error. The degrees of freedom for the 1,2,3,4, 5 years of horizons were 1295, 1283, 1271, 1259 and 1247 respectively.

Period	Paired Correlation	Mean Differences	Std. error	t-statistic
1 year	0.368	0.01	0.0007	14.28
2 year	0.329	0.05	0.0014	28.57
3 year	0.297	0.01	0.0022	4.54
4 year	0.268	0.02	0.0031	6.45
5 year	0.247	0.03	0.0041	7.31

The Buy and Hold significantly outperforms the strategy. The

mean difference of zero is rejected for all horizons. As the horizon increases, the t-statistic becomes smaller. This means that the difference in volatility between Buy and Hold and the strategy becomes larger over higher horizons. However, there is no uniformity in the magnitude of the t-statistics over the different horizons. For example, the t-statistic increases from the first year to the second, and afterwards goes down in the third year.

Test of Riskiness

As a further test, the Sweeney X is computed to take into account the number of periods that the asset is stored in the equity or in the bond market. When the asset is held in the stock market, it possesses higher risk than when the asset is kept in the bond market. The X statistic turns out to be 0.0022 with a standard error of 5.4×10^{-5} . The t-statistic rejects the null hypothesis of equal returns for both the rule and the Buy and Hold strategy after accounting for the riskiness of the equity index.

Sampling with Replacement

The earlier results for the different horizons are based on the whole sample of returns. The results may be biased due to periods of exceptionally high returns. In that case, the trading strategy's performance may be simply due to 'luck'. In the present case, it could be mainly due to bad luck since it missed on high returns in the equity market. In the following test, random dates are picked from the larger sample, and the success of the rule against Buy and Hold strategy are investigated. The sampling exercise is similar to the one explained in the first chapter. A vector of dates is generated using a random number generator where a vector containing ele-

ments between zero and one are randomly chosen from the uniform probability distribution. The results are reported in table 3.5.

Table 3.5:

Replacement Sampling.

The figures illustrate the percentage number of times the trading rule strictly beats the Buy and Hold strategy.

Period	20	40	80	120
1 year	40	40	46.25	43.13
2 year	35	47.5	43.75	43.75
3 year	35	47.5	40	40.63
4 year	40	47.5	42.5	41.88
5 year	45	52.5	41.25	41.88

The strategy does not outperform the Buy and Hold most of the times. This is shown by the fact that the percentage is lower than 50 % over the sample adopted. It can be seen that for a higher number of dates being selected in the sample, the strategy is still being outperformed by Buy and Hold, irrespective of the time frame the strategy is put to use. Compared to the results presented in chapter two, which showed enormous success of the trading strategy, the trading rule in the state space model fares worse because of the higher discount rate. This is shown in figure 3.4.

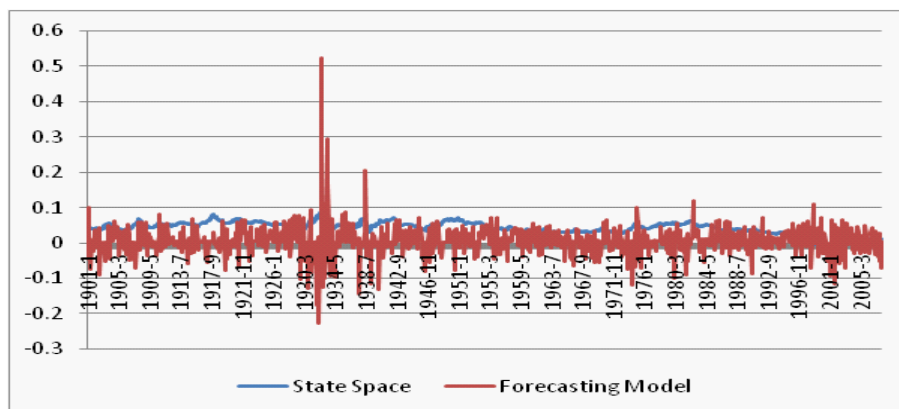


Figure 3.4: **Plot of Discount Rate.** The figure shows the discount rate from the state space model and the forecast model B from chapter two.

The figure shows that the discount rate under the state space model is significantly higher throughout the sample. There are periods when the forecast is higher, such as the late 90's onwards, but this may be attributed to the volatile nature of the series. In such a case, it is obvious that due to the inflated discount rate, the present value will be lower. The strategy therefore postulates going on bonds which yield a lower return. Despite being more volatile, the discount rate appears

3.5 Conclusion

The contribution of this chapter is twofold. It retrieves time series of expected returns and dividend growth from a structural present value. The structural present value model starts with the Campbell-Shiller identity which relates the price dividend ratio to the expected returns and expected dividend growth rate. In order to ensure that the model contains two models for two unobservables, realized dividend growth is also used as a measurement variable, and is directly linked to the expected dividend growth. The model is applied to

monthly data from January 1900 to December 2008. The results show that both expected dividend growth rate and expected returns are highly persistent, and close to unit root in the case of returns.

The Kalman filter yields a smooth filtered forecast of expected returns and expected dividend growth which is used in the present value formula. This may be contrasted to the highly volatile series of average realized returns. The present value is used to identify whether equity market is underpriced. The strategy performs poorly as it does not identify bull states properly. The strategy fails because the discount rate is high, which leads to the present value being lower more often than it should be. Hence it misses out on bullish trends, and advises to hold bonds mostly. In a nutshell it is possible that simple forecast models of returns may have a better appeal for the present value strategy. An attempt to implement the model on an annual basis proved to be futile and was deemed not interesting to be reported. In that case, the strategy postulated going on bonds even in both periods of boom and crashes.

The return on the strategy is on average 1 % lower than the Buy and Hold although the terminal wealth is higher. This is simply due to correct switches in the later years. Three robustness tests were considered namely the Sharpe Ratio, Replacement Sampling and Sweeney Statistic. The tests all fared poorly. This chapter contributes in showing an alternative way to compute expected returns and expected dividend growth. The Kalman Filter has not been applied in the context of a trading strategy. The work may be extended to test whether the strategy with the new discount rates work better with individual securities, or in other stock markets.

Chapter 4

Predictive Ability of Expected Returns

4.1 Introduction

Over the years, the return predictability literature has demonstrated that some financial and macroeconomic indicators may forecast returns. Some examples are the price-dividend ratio, price-earnings ratio and consumption to wealth ratio. These indicators capture cyclical tendencies in macroeconomic variables. One potential predictor of realized returns may be lagged expected returns. The previous chapter showed how returns may be derived from a structural present value model. In this chapter, the derived time series is exploited for simple predictability purposes¹. Realized returns are a weighted function of expected returns under fundamentals and speculation. The expected returns are derived using the state space present value methodology explained in chapter three. The study covers both the in-sample and out-of-sample predictability of returns using simple Ordinary Least Squares and Vector Autoregression models.

¹The objective of the study is to compare the performance of expected returns and the price-dividend ratio.

The expected returns and expected dividend growth series are filtered from realized observations based on the Kalman procedure, where the values for the expected variable are updated as a new observation of the realized value enters the information set. The law of motion for the price-dividend ratio is derived assuming that the expected returns and the dividend growth rate follow a stationary autoregressive process of order one². The state space model is derived from the present value relationship among price-dividend ratio, expected returns and expected dividend growth. The Kalman Filter is applied to the model parameters which are optimized using the Maximum Likelihood procedure. The model is estimated using annual data from 1900 to 2008.

The first part of the paper reviews the literature on predictability. A brief summary of the structural model estimation already outlined in detail in chapter 3 follows. The second part of the empirical analysis delves in the predictive ability of the two different variables. Actual returns are regressed on the past filtered observations of expected returns. We assess the in-sample predictability based on the goodness of fit measure and the out-of-sample accuracy based on the mean squared error. We also model the returns in a multivariate framework. In this setting, a VAR model is constructed using both actual and expected series for the variable of interest.

4.2 Literature review

The literature on return and dividend growth predictability has highlighted many variables which have reasonable forecasting power both in-sample and out-of-sample. For example, earnings to price

²If the latent variables do not follow an AR(1), then the results may show spurious predictability.

ratio, book to market ratio, consumption to wealth ratio, labour to consumption ratio, housing collateral ratio, or the cross section of risk. For an interesting literature see Lamont(1998), Baker and Wurgler (2000), Lettau and Ludvigson (2001), Menzly, Santos, and Veronesi (2004), Lustig and Van Nieuwerburgh (2005), Piazzesi, Schneider and Tuzel (2006), Polk, Thompson and Vuolteenaho (2006), Cochrane (2008), Koijen and Van Binsbergen (2011). Moreover, there is also an interesting literature concentrating on the econometric properties of predictive regression models. This literature involves issues such as time varying parameters, out-of-sample forecasts, structural breaks and weak exogeneity of predictors. Interesting papers in this field include Bossaerts and Hillion (1999), Stambaugh (1999), Campbell and Yogo(2006), Ferson, Sarkisson and Simin (2003), Valkanov(2003), Goyal and Welch (2003 and 2007), Inoue and Killian (2004) Lewellen (2004), Ahmihud and Hurvich (2004) and Ang and Bekaert (2006).

The central tenet of the present paper is to see whether returns may be predicted from the expected value. The price-dividend ratio, in itself a predictor, is the key measurement variable which enables the separation of the expected returns and dividend growth. The price-dividend ratio inherits the interesting property of being in line with the present value framework assuming rational expectations. This is proved by Campbell and Shiller (1988) who show that the price-dividend ratio is an approximate function of returns and growth. Papers discussing predictability of returns and dividend growth using this approach include Fama and French (1988), Campbell and Shiller (1991), Timmermann (1993, 1996), Lewellen (2004), Cochrane (2008), Koijen and Van Binsbergen (2010). It can be shown from the Campbell and Shiller (1988) log linearized form that as long as the expected returns and dividend growth process are stationary, deviations of the price-dividend ratio from its mean

may either predict returns or dividend growth³.

4.3 Methodology

In this section, a brief summary of the present value approach explained in Chapter 2 is presented. The corresponding predictive regressions are also presented.

4.3.1 State Space Representation

The returns on stocks, price-dividend ratio and the dividend growth are given by the following equations.

$$r_{t+1} = \log\left(\frac{P_{t+1} + D_{t+1}}{P_t}\right). \quad (4.1)$$

$$PD_t = \frac{P_t}{D_t}. \quad (4.2)$$

$$\Delta d_{t+1} = \log\left(\frac{D_{t+1}}{D_t}\right). \quad (4.3)$$

The autoregressive demeaned expected returns and expected dividend growth are written as (4.4) and (4.5) respectively:

$$\mu_{t+1} - \delta_0 = \delta_1(\mu_t - \delta_0) + \varepsilon_{t+1}^\mu. \quad (4.4)$$

$$g_{t+1} - \gamma_0 = \gamma_1(g_t - \gamma_0) + \varepsilon_{t+1}^g. \quad (4.5)$$

To prevent any measure of underidentifiability, the realized dividend growth rate is defined as the expected dividend growth rate and the unobserved shock ε_{t+1}^d , where by :

³This implies that the decomposed expected returns should have a lower variance and hence it may provide better point forecasts than the price dividend ratio.

$$\Delta d_{t+1} = g_t + \varepsilon_{t+1}^d. \quad (4.6)$$

To derive the state space model, the Campbell and Shiller (1988) log linearized return approximation may be written as :

$$r_{t+1} = \kappa + \rho p d_{t+1} + \Delta d_{t+1} - p d_t. \quad (4.7)$$

The reduced form model may be written as:

$$\tilde{g}_{t+1} = \gamma_1 \tilde{g}_t + \varepsilon_{t+1}^g, \quad (4.8)$$

$$\Delta d_{t+1} = \gamma_0 + \tilde{g}_t + \varepsilon_{t+1}^d, \quad (4.9)$$

$$p d_{t+1} = (1 - \delta_1) A + B_2 (\gamma_1 - \delta_1) \tilde{g}_t + \delta_1 p d_t - B_1 \varepsilon_{t+1}^\mu + B_2 \varepsilon_{t+1}^g, \quad (4.10)$$

where $A = \frac{\kappa}{1-\rho} + \frac{\gamma_0 - \delta_0}{1-\rho}$, $B_1 = \frac{1}{1-\rho\delta_1}$, $B_2 = \frac{1}{1-\rho\gamma_1}$.

(4.8) is the transition (state) equation. The measurement equation relates the observable variable to the unobserved variables. In our case this is given by (4.9) and (4.10). The Kalman Filter is implemented to obtain the likelihood, which is then maximized over the following vector of parameters.

$$\Theta = (\gamma_0, \delta_0, \gamma_1, \delta_1, \sigma_g, \sigma_\mu, \sigma_d, \rho_{g\mu}, \rho_{gd}, \rho_{\mu d}) \quad (4.11)$$

The parameters are estimated over the whole sample size. In other words, the present value parameters are fixed throughout the forecasts. It should be stressed that the expected returns and expected dividend growth are still time varying.

4.3.2 Predictive Accuracy

In this section, predictive regressions are explained. Predictability implies that a variable may be forecast from other factors based on the information set available at that time. In the univariate setting, realized returns is regressed on the first lag of expected returns and the price-dividend ratio. (4.12) and (4.13) illustrate the predictive regression for realized returns by the lagged expected returns and price-dividend ratio.

$$r_t = \beta_0 + \beta_1 \mu_{t-1} + \varepsilon_t, \quad (4.12)$$

$$r_t = \theta_0 + \theta_1 pd_{t-1} + v_t, \quad (4.13)$$

where ε_t and v_t are the disturbances.

Higher order lags may be included, but at the cost of losing degrees of freedom. Given that this study looks at predictability based on annual data, the sample size is relatively small. Therefore forecast accuracy measures may suffer from small sample bias. However, we shall attempt to increase the number of lags in the VAR framework.

The Vector Autoregression model is intuitively interesting to investigate in the predictability literature given that agents may update their expectations of returns given the realized value. It is possible that the realized returns in a lag period may influence expected returns or the price-dividend ratio in the next. In other words, there are feedback effects among the different predictors. A bivariate vector autoregression is put forward to account for such a possibility:

$$Y_t = C + \sum_{i=1}^p A_i Y_{t-i} + u_t \quad (4.14)$$

We use this generic form for both expected returns and the price-dividend ratio. In the case of expected returns, $Y_t = [r_t \ \mu_t]'$. A is a 2×2 matrix of coefficients and for p lags, matrix. p is usually the VAR order and shall be set to 1, 2 and 3. u_t is a vector of disturbances. C is a vector of intercept terms.

Cumulative (Horizon) Returns

If the predictor variable is stationary but highly persistent, predictability will be better for higher horizon cumulated returns. For instance, predictability (both in-sample and out-of-sample) of returns cumulated over three years should be higher than for a single year⁴. In this case the cumulated returns over j periods is written as $\sum_{j=1}^k r_{t+j}$. The latter is the simple arithmetic compounding of log returns. We shall empirically test whether the predictability of the cumulated returns from the expected returns and the price-dividend ratio. To illustrate this fact, consider the simple linear regression model(4.16):

$$r_{t+1} = \beta x_t + \varepsilon_{t+1}, \quad (4.15)$$

$$x_t = \gamma x_{t-1} + \varepsilon_t^x. \quad (4.16)$$

r_{t+1} is predicted from x_t . However, x_t is modeled by an autoregressive process where γ captures the persistence parameter. The return forecast for the second year is predicted by the predicted variable from $t + 1$.

⁴An excellent explanation on how persistence creates higher predictability is available in Chapter 20 in Asset Pricing from Cochrane (2005).

$$r_{t+2} = \beta x_{t+1} + \varepsilon_{t+1}.$$

$\sum_{j=1}^2 r_{t+j}$ can therefore be written as

$$\begin{aligned} r_{t+1} + r_{t+2} &= \beta x_t + \beta(\gamma x_t + \varepsilon_{t+1}^x) + \varepsilon_{t+1} + \varepsilon_{t+2}, \\ &= \beta(1 + \gamma)x_t + \beta\varepsilon_{t+1}^x + \varepsilon_{t+1} + \varepsilon_{t+2}. \end{aligned}$$

The generalization to k periods can therefore be written as;

$$\sum_{j=1}^k r_{t+j} = (1 + \gamma + \gamma^2 + \dots + \gamma^{k-1})x_t + \epsilon_t$$

where ϵ_t is an error term which is made up of the sum of individual

error terms $(\varepsilon_{t+1}, \dots, \varepsilon_{t+k})$. Therefore the standard OLS coefficient will rise as k increases if γ is close to one. It is obvious that the R^2 will rise as well.

4.3.3 Stambaugh Bias

One potential problem highlighted by Stambaugh (1999) is the bias occurring in the presence of persistent regressors in univariate regressions. It is worth mentioning that it is not a major issue in the present study as both the price-dividend ratio and the expected returns share the same persistence level. According to the theory on predictability, there should be an upward bias in the predictive parameter which occurs because of the negative correlation between the innovation terms of return and the predictor variable and also the high persistence in the predictor variable.

$$\begin{aligned} r_{t+1} &= \alpha + \beta x_t + \varepsilon_{y,t}, \\ x_t &= \theta + \gamma x_{t-1} + \varepsilon_{x,t}. \end{aligned} \tag{4.17}$$

From (4.17), the Stambaugh bias can be quantified as (4.18):

$$E(\widehat{\beta} - \beta) = \frac{\sigma_{yx}}{\sigma_x^2} E(\widehat{\gamma} - \gamma). \tag{4.18}$$

In the presence of this bias, the coefficient should be corrected according to the persistence level or assuming a unit root. For interesting discussions, see Stambaugh (1999), Ahmihud and Hurvich (2004) and Lewellen (2004). γ , the persistence factor is 0.951 for the price-dividend ratio and 0.947 for the expected returns. Hence, both persistence factors are roughly the same and hence comparing biased predictor coefficients who share the same level of persistence may not matter much for this empirical study.⁵

4.4 Results

We first report the results of the optimization of the state space model defined by (4.8), (4.9) and (4.10). The model was estimated for the annual data 1900-2008 on the S&P 500 index. The result from the state space optimization is shown in table 4.1.

⁵Moreover, inference on the coefficient is not the main question being asked here.

Table 4.1:

Estimation of State Space Parameters.

The table shows the estimated parameters and standard errors from the optimization exercise using data from 1900-2008. The standard errors are again computed from the Hessian Matrix.

Parameter	Coefficient	Std error
γ_0	0.014	0.011
δ_0	0.052	0.016
γ_1	0.072	0.199
δ_1	0.946	0.040
σ_g	0.093	0.058
σ_d	0.014	0.006
σ_μ	0.069	0.077
$\rho_{g\mu}$	0.480	0.471
$\rho_{\mu D}$	-0.381	0.392

The respective unconditional mean of expected returns and dividend growth rate are 5.2 % and 1.4 %. Low variation around the parameter estimates is observed. Interestingly, the low autoregressive parameter of growth implies that expected dividend growth rate in itself cannot be predicted from past lags, further reinforcing previous empirical evidence that dividend growth is unpredictable. However, the interesting result comes from the autoregressive coefficient of the expected returns which shows high persistence. The high autoregressive coefficient implies that a shock to expected returns may take time to disappear. This contrasts the findings of realized return exhibiting on serial correlation in the first moments.

4.4.1 Statistical Properties of Expected Returns and Growth

In this section, the statistical features of the derived expected returns and dividend growth are described. The summary statistics for the series are given in table A.18 in the appendix. Stationarity

and long memory tests are reported in table A.19. Both μ_t and g_t have unconditional mean properties close to the realized values. The filtered variables possess a lower volatility than the observed values. Expectations in general are meant to be less volatile than realized values. The third moments illustrate a higher negative skewness value, illustrating that the variance in expected returns is explained by extreme deviations from the mean.

The null hypothesis of stationarity is not rejected in most cases. The nonstationarity test results show that there is high evidence of stationarity in realized returns, and also realized and expected dividend growth. However the expected returns is purely nonstationary according to the various tests. In the present setting, nonstationarity is not a problem as long as the root is non-explosive⁶. The results point out that expected returns may follow a random walk. We also report the cross correlations between the expected and realized series in tables A.20 and A.10. The correlation of the returns with its own past lags is quite low, although a relatively strong mean reversion is witnessed at lag 2. The expected returns has strong positive correlation with its past lagged values. The actual returns is weakly positively correlated with the lags of the expected returns. Expected returns on the other hand appear to be weakly negatively correlated with past lags of actual returns.

4.4.2 Predictive Accuracy

In this section, the results on the in-sample predictive accuracy are presented. This involves estimating (4.12) and (4.13) where actual returns are predicted from expected returns and the price-dividend ratio respectively and comparing the R-squared. We look at return predictability over 5 years.

⁶A series may be nonstationary but non-explosive. In the present setting, a near unit root

Table 4.2:

In-sample Predictability of Expected Returns.

The first column shows the number of years of accumulated returns in-sample. The last column reflects in sample predictability over the horizon. The standard errors are computed using the robust formulae.

k	β_1	Std error	T-ratio	R-Squared
1	0.981	0.457	2.15	0.040
2	1.46	0.645	2.26	0.042
3	2.102	0.760	2.77	0.063
4	2.844	0.880	3.23	0.084
5	3.399	0.961	3.54	0.096

Table 4.3:

In-sample Predictability of Price-dividend ratio.

The first column shows the number of years of accumulated returns in-sample. The last column reflects in sample predictability over the horizon. The standard errors are computed using the robust formulae.

	θ_1	Std error	T-ratio	R-Squared
1	-0.0887	0.042	-2.103	0.038
2	-0.1345	0.059	-2.264	0.042
3	-0.1957	0.069	-2.809	0.064
4	-0.2618	0.081	-3.242	0.084
5	-0.3107	0.088	-3.532	0.094

Tables 4.2 and 4.3 show that both expected returns and price-dividend ratio share the same level of predictive accuracy. This can easily be seen from the given R-squared. The expected returns marginally overcomes the price-dividend ratio as a predictor for the one and five year horizon. The reason as to why these two predictors predict as accurately as each other may be because dividend growth does not contribute to returns predictability directly. Using the Campbell and Shiller approximation, almost all of the variation

behaviour is witnessed since the autoregressive coefficient is less than one.

in the price-dividend ratio comes from the movement in expected returns. The predictive power of both models tend to improve over time, as different returns horizon are taken into account. The long run predictability is improved since both expected returns and the price-dividend ratio are highly persistent. Their persistence ensures that predictability power is accumulated over time. In simple terms, the predictive power for 2 years includes the predictive power for one year as well.

4.4.3 Out-of-Sample Forecast

In this section, the out-of-sample forecast predictive accuracy is analyzed. In the case of the out-of-sample forecast, we estimate the parameters from the present value model (4.11) until the year 2000 and forecast out-of-sample using (4.12) and (4.13) for 8 steps ahead. The mean squared error is computed and reported in tables 4.4 and 4.5 for μ_{t-1} and pd_{t-1} . The root mean squared error are also plotted in figures 4.1 and 4.2.

Table 4.4:

Out-of-sample Mean Squared Error for Expected Returns.

The table illustrates the out-of-sample predictability over the period 2001-2008 for the different horizon cumulated returns from 1 to 5 years when returns are predicted by the filtered returns series.* represents the horizon when the mean squared error is highest ** represents the lowest mean squared error.

Forecast Horizon	1 Year	2 Year	3 Year	4 Year	5 Year
1	0.00986	0.17913*	0.04721*	0.00121	0.00464
2	0.02226*	0.06570	0.00965	0.00266	0.03765
3	0.00065	0.00833	0.00075	0.03553*	0.09522*
4	0.00004**	0.00291	0.00368	0.03819	0.02861
5	0.00554	0.00527	0.01741	0.00384	0.07713
6	0.00109	0.00009**	0.00017**	0.00009**	0.00001**
7	0.00103	0.00716	0.00803	0.00432	0.00475
8	0.00103	0.02050	0.02215	0.01527	0.01522

Table 4.5:

Out-of-sample Mean Squared Error for Price-dividend ratio.

The table shows the out-of-sample mean squared error over the period 2001-2008 when returns are predicted by pd_{t-1} .

Forecast Horizon	1 Year	2 Year	3 Year	4 Year	5 Year
1	0.05945	0.24030*	0.04674*	0.00144	0.00331
2	0.08510*	0.10404	0.00959	0.00221	0.03303
3	0.02627	0.02403	0.00071	0.03323	0.08657
4	0.01417	0.01265	0.00352	0.03557*	0.02383
5	0.00046**	0.00074**	0.01641	0.00521	0.08753*
6	0.00691	0.00405	0.00022**	0.00001**	0.00014**
7	0.00639	0.01874	0.00826	0.00518	0.00677
8	0.01998	0.03771	0.02250	0.01678	0.01858

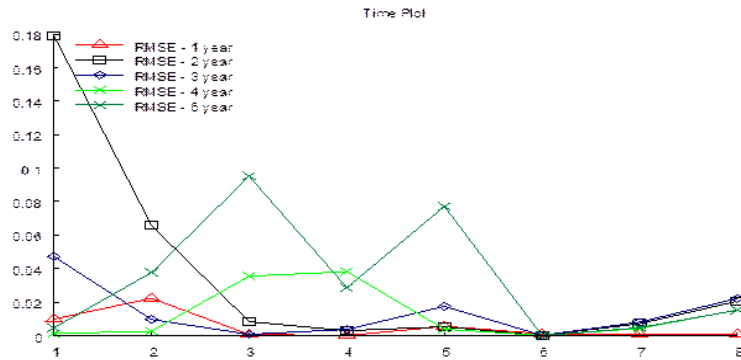


Figure 4.1: **Plot of root mean squared error for expected returns.** The figure shows the plot of the root mean squared error for each of the accumulated returns horizon over the period 2001-2008. The predictor variable in this case is expected returns.

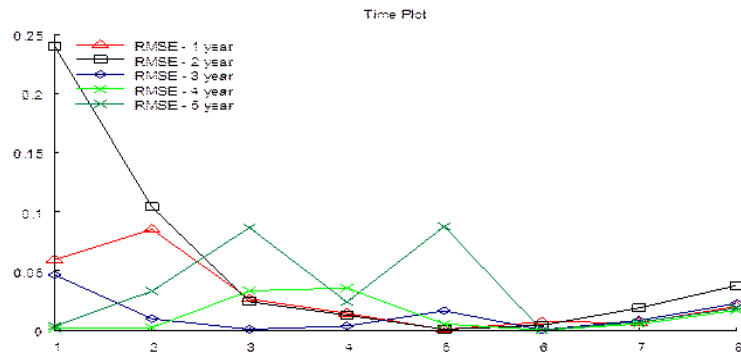


Figure 4.2: **Plot of Root Mean Squared Error for Price-dividend ratio.** The figure shows the plot of the root mean squared error for each of the accumulated returns horizon over the period 2001-2008. The predictor variable in this case is the price-dividend ratio.

The predictive ability through μ_{t-1} is relatively low over short horizons. Generally the model tends to fare worse during the first three years of forecasting. For the four and five years' cumulative returns, the worst predictability is in the third year and the best predictability is achieved in year 6. The graphical plots 4.1 and 4.2 show that both predictors are relatively good out-of-sample over long horizons. The root mean squared error tends to decrease over time. The only exception is the 5 year of accumulated returns. This may be accounted for by structural breaks prior to 2000 and the small sample bias. When returns are cumulated over five years, the estimation loses four observations, which implies that the estimates may be less consistent.

The model fares badly in the first year. As the horizon increases, the mean squared error decreases for both the price-dividend ratio and expected returns. The best prediction horizon is four and five years. For the first two years of accumulated returns, the best horizon is year 5. For the remaining three years, the best out-of-sample predictability is the 6 year horizon. A general trend depicted is that for one and two year returns, expected returns is the best predictor of returns. However, the best predictor for the long run over the whole 8 year horizons tend to be the price-dividend ratio. This may be explained through the fact that dividend growth may add further information which is captured in the price-dividend ratio but in expected returns.

4.4.4 VAR models

In this section, the results from the vector autoregression for three different lag orders are reported. The predictive regression results from both the expected return and the price-dividend ratio are reported in tables 4.6 and 4.7.

Table 4.6:

**Results from VAR model with Realized and Expected returns:
Sample 1900-2008.**

The table shows the parameters and standard errors from the VAR. P refers to the number of lags in the VAR model. ** statistical significance at the 1 % level; * denotes significance at the 5 % level. The figures inside the brackets refer to the p-values.

	P =1		P =2		P =3	
	\mathbf{R}_t	μ_{t-1}	\mathbf{R}_t	μ_{t-1}	\mathbf{R}_t	μ_{t-1}
C	0.056 (0.021)*	0.0596 (0.002)	0.0552 (0)**	0.059 (0.002)	0.0453 (0.018)*	0.051 (0.036)
\mathbf{R}_{t-1}	0.042 (0.691)	0.0003 (0.977)	0.075 (0.47)	0.004 (0.432)	0.075 (0.48)	0.003 (0.442)
μ_{t-1}	0.594 (0.215)	0.9201 (0)**	-0.361 (0.706)	0.942 (0)**	0.513 (0.736)	1.079 (0)**
\mathbf{R}_{t-2}			-0.2310 (0.026)	-0.065 (0)**	-0.226 (0.049)	-0.064 (0)**
μ_{t-2}			1.3022 (0.203)	0.0387 (0.616)	0.127 (0.952)	-0.112 (0.334)
\mathbf{R}_{t-3}					0.109 (0.488)	0.0199 (0.046)
μ_{t-3}					0.356 (0.786)	0.026 (0.728)
Adj R- Squared	0.018	0.839	0.079	0.935	0.086	0.950
Akaike	422.47		472.473		477.63	
Schwartz	430.54		459.062		458.92	

Table 4.7:

Results from VAR model with Realized returns and Price-dividend ratio: Sample 1900-2008.

The table shows the parameters and standard errors from the VAR. P refers to the number of lags in the VAR model. ** statistical significance at the 1 % level; * denotes significance at the 5 % level. The figures inside the brackets refer to the p-values.

	P =1		P =2		P =3	
	R_t	PD_{t-1}	R_t	PD_{t-1}	R_t	PD_{t-1}
C	0.0517 (0.015)**	3.217 (0)**	0.049 (0)**	3.235 (0)**	0.048 (0.001)**	3.239 (0)**
R_{t-1}	0.05274 (0.617)	-0.276 (0.17)	0.054 (0.609)	-0.054 (0.272)	0.078 (0.464)	-0.052 (0.276)
PD_{t-1}	0.05966 (0.039)	0.551 (0.027)*	-0.044 (0.331)	0.898 (0)**	-0.078 (0.649)	1.108 (0)**
R_{t-2}			-0.219 (0.034)*	0.702 (0)**	-0.234 (0.036)*	0.707 (0)**
PD_{t-2}			0.042 (0.105)	0.098 (0)**	-0.015 (0.928)	-0.121 (0.326)
R_{t-3}					0.133 (0.44)	-0.149 (0.25)
PD_{t-3}					0.004 (0.903)	0.001 (0.975)
Adj R-Squared	0.0299	0.4861	0.0764	0.952	0.086	0.954
Akaike	106.257		230.792		227.592	
Schwartz	98.1831		217.381		208.882	

The individual significance (as measured by the t-ratio) for the predictor variables tend to be different. In the case of expected returns, the only variable which matters is the second lag of realized returns.⁷ Among the three different lags, we find that predictability is best achieved by looking at a model with 2 lags. The Adjusted R-squared is 0.08. For expected returns, the optimal lag order is given by P=2 (Schwartz criterion) and P=3 (Akaike). However, the

⁷In the case of multivariate models, the Stambaugh bias cannot be signed. See Stambaugh (1999)

optimal number of lags for the price-dividend ratio is 2 in both cases.

Dynamic out-of-sample forecasts using the VAR model are also reported. The corresponding mean squared error is reported in tables 4.8 and 4.9 for the expected returns and price-dividend ratio respectively.

Table 4.8:

Recursive Mean Squared Error for Expected Returns.

The predictor variable is μ_{t-1} . * illustrates the largest recursive mean squared error for each horizon. ** illustrates the lowest mean squared error. P relates to the order of the VAR model.

Year	1 period Return			2 period Return			3 period Return			4 period Return		
	P = 1	P = 2	P = 3	P = 1	P = 2	P = 3	P = 1	P = 2	P = 3	P = 1	P = 2	P = 3
1	0.0396	0.0309	0.0274	0.0449*	0.0475*	0.0488*	0.0308*	0.0156*	0.0075*	0.005*	0.004	0.003
2	0.0602*	0.0519*	0.0473*	0.0400	0.0294	0.0215	0.0116	0.0054	0.0023	0.001	0.001	0.000
3	0.0133	0.0095	0.0071	0.0045	0.0020	0.0007	0.0000	0.0008	0.0020	0.003	0.004*	0.005*
4	0.0061	0.0041	0.0025	0.0019	0.0007	0.0001**	0.0001**	0.0009	0.0016	0.002	0.003	0.003
5	0.0046	0.0021	0.0008	0.0006	0.0000**	0.0002	0.0007	0.0020	0.0027	0.002	0.001	0.001
6	0.0022	0.0006	0.0000	0.0001**	0.0001	0.0006	0.0005	0.0008	0.0007	0.000	0.000**	0.000**
7	0.0021	0.0005**	0.0000**	0.0008	0.0004	0.0003	0.0004	0.0001**	0.0001**	0.000**	0.000	0.000
8	0.0008	0.0004	0.0008	0.0005	0.0003	0.0003	0.0002	0.0002	0.0002	0.000	0.000	0.000

Table 4.9:

Recursive Mean Squared Error for Price-dividend ratio.

The table shows the root mean squared error for the different levels of cumulated returns when the predictor is pd_{t-1} . * illustrates the largest mean squared error. ** illustrates the lowest mean squared error.

Year	1 period Return			2 period Return			3 period Return			4 period Return		
	$P = 1$	$P = 2$	$P = 3$	$P = 1$	$P = 2$	$P = 3$	$P = 1$	$P = 2$	$P = 3$	$P = 1$	$P = 2$	$P = 3$
1	0.026	0.023	0.021	0.033*	0.036*	0.038*	0.023*	0.011*	0.005*	0.003	0.002	0.001
2	0.042*	0.041*	0.040*	0.029	0.021	0.015	0.007	0.003	0.001	0.001	0.000**	0.000**
3	0.005	0.006	0.005	0.001	0.000	0.000	0.000	0.001	0.003	0.004*	0.005*	0.006*
4	0.001	0.002	0.001	0.000	0.000	0.000**	0.000	0.001	0.002	0.002	0.003	0.003
5	0.001	0.001	0.000	0.000**	0.000**	0.000	0.000	0.001	0.003	0.002	0.001	0.001
6	0.000**	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.001	0.000	0.000	0.000
7	0.000	0.000**	0.000**	0.000	0.000	0.000	0.000**	0.000**	0.000**	0.000	0.000	0.000
8	0.006	0.005	0.004	0.004	0.004	0.003	0.003	0.003	0.002	0.003	0.003	0.002

The forecasts are unlikely to reproduce the current actual returns. They do not follow the same direction. If the forecast is showing a positive return, the realized return is negative. Moreover, forecast returns tend to be more or less stable where as the actual returns tend to fluctuate a lot. However, the forecasts for the expected returns (fitted values) tend to reproduce the expected returns (actual values). For the expected returns, the VAR(1) and VAR(3) are better for 2004 and 2005 respectively. After 2005, the VAR(2) appears to be the best model for forecasting expected returns.

4.5 Economic Value of Return Predictability

In this section we delve into the economic value of the return predictability. We use the formula of Taylor(2012). The statistic used refers to the maximum amount that an individual is willing to pay for predictability knowledge. The formula is as follows:

$$\delta_H = \frac{1}{2\theta} \ln\left[\frac{1}{1 - R_H^2}\right] \quad (4.19)$$

where R_H^2 is the goodness of fit from a regression of H-period horizon on the predictor variable, and δ_H is the maximum amount an H -period horizon uninformed investor is willing to pay for the use of the informed investor's conditional information. θ is the Arrow-Pratt relative risk aversion.⁸ We selectively report cases for insample predictability in table (4.10).

The table shows that both expected returns and price-dividend ratio have the same performance fees. At the 5-horizon, δ_H is slightly higher for expected returns. In the case of the price-dividend ratio,

⁸ θ may take different values. In Taylor(2012), θ depends on the portfolio weights of the risky and risk-free asset.

Table 4.10:

Economic Value of Predictability.

The table reports the formula from (4.19). The cases reported are the univariate 1 and 5 year horizon for the univariate case and the three lags for the VAR case. θ is set to 4 which corresponds to a weight of 0.5 in the equity index.

Model	Predictor variable	
	μ_{t-1}	\mathbf{PD}_{t-1}
Univariate Regression:		
1 – horizon	0.005	0.005
5 – horizon	0.013	0.012
VAR Models		
$P = 1$	0.002	0.004
$P = 2$	0.010	0.010
$P = 3$	0.011	0.011

δ_H is relatively the same for both predictors. However it is interesting to note that there is significant difference for the two lags of the VAR model. The performance fees do not differ for $P = 2$ and 3. It is interesting to note that while there is a higher statistical difference than economic significance.

4.6 Conclusion

The conclusions of this paper rest on some important assumptions for deriving the expected returns and expected dividend growth rate. Firstly, expected returns is modeled in the context of the simple present value. Secondly, expected returns and dividend growth rate are approximated by a stationary autoregressive process of order one. The findings show that expected returns does not improve on the price-dividend ratio in predicting both in-sample and out-of-sample. If the state equations are improperly specified, then the expected returns may not capture true expectations over time.

Empirically, we found evidence of weak predictability in both the OLS and VAR models for expected returns. The Ordinary Least Squares was adopted for both insample and out of sample predictability. Cumulated returns up to 5 years were considered as the dependent variable. The R-squared was used as the measure of in-sample performance. The out-of-sample consisted of 8 year horizon, and the performance was evaluated using the Mean Squared Error. In the case of the VAR, the focus was on the number of lags to be adopted. Again the performance was judged using the R-squared and the Mean Squared Error. The economic significance was also assessed using Taylor's (2012) formula.

The VAR produced better forecasts both in-sample and out-of-sample for the period 2001- 2008. Weak predictability was witnessed the maximum R-squared from the different models tend to be 0.09. In the OLS models, it is found that the R-squared tends to increase with the horizon for both regressors. In the case of the out-of sample results, the model appears to fare better in the case of 5-6 horizons. The same finding was witnessed with the VAR model. The optimal VAR lag order was found to be 2 in both cases. Our results seem to confirm the previous results from Cochrane (2009) where similar predictability levels was found both insample and out-of-sample.

The main contribution of this chapter is the comparison of the price-dividend ratio with the expected returns, which according to the state space model is a less noisy predictor since it removes the unpredictable component of dividend growth. Although this is an actively researched area, there are no studies which consider filtered expected returns forecast as a predictor of realized returns. The results of this chapter might be of interest for it shows that the price-dividend ratio is still a better predictor variable. Hence, it may be

implied that dividend growth may have some marginal explanatory power for returns.

Chapter 5

Modeling the Persistence in Expected Returns

5.1 Introduction

The previous two chapters showed that expected returns are persistent over time. The autoregressive coefficient in the expected returns specification is high and near unity. Kojen and Van Binsbergen (2010), hence KVB, also found a similar result when expected returns are derived from the present value.¹ This implies that the autocorrelation at higher lags are different from zero. Persistence is defined by how highly a series is serially correlated over time. In this chapter, I model the persistence in expected returns using two models with fractionally integrated processes. The persistence in the expected returns is modelled explicitly by assuming that the series follow an autoregressive fractionally integrated process (ARFIMA(p, d, q)), where p is the number autoregressive lags, d is the memory parameter and q is the number of moving average lags. This process accounts for the possibility of a series

¹This study found that the autoregressive parameter is 0.932 with a standard error of 0.128. Kalman Filter estimates are still consistent in the presence of unit roots (See Brockwell and Davis 1991).

having long memory, where the persistence or memory is measured by ‘ d ’. In empirical applications, many series such as returns volatility and exchange rates are found to exhibit long memory. To the best of our knowledge, no study has yet adapted fractional and long memory models to the present value.

ARFIMA(p, d, q) models are specifications of processes where a fractional order of integration is involved, usually defined by the ‘ d ’ parameter. A statistically significant d may imply high persistence (long memory) or anti-persistence, depending on the range where d lies. The process makes the distinction between short (p and q) and the long range (d) components. When $0 < d < 0.5$, the process exhibits long memory, and is stationary. When $d > 0.5$, the process is long memory but nonstationary as the sum of variances of such process go to infinity. By differencing the process d times, the process becomes a stationary ($I(0)$) process, with short memory. An important advantage of this transformation is that it provides stronger consistency results for tests on the stationary series. An introduction to such models is available in Beran (1994).

Two structural models are considered. In the first case, expected returns are modeled as an ARFIMA process within the Kalman Filter as a property of one of the transition equations. This approach is similar to the one adopted in the previous chapters where expected returns are expected to follow an autoregressive process. In the second case, parametric specifications of the dividend growth process are used to proxy expectations with forecasts, similar to chapter two. The forecasts are then replaced in the present value to yield a series for expected returns. The ARFIMA model has a better appeal in terms of the economic intuition guiding the process. Based on the seminal work of Granger and Joyeux (1980), generated long memory in time series data occurs through aggregation of

micro-processes. These processes are based on difference equations in dynamic economic models. It can be shown that, when individual processes (for instance expected returns) are aggregated, then the aggregated process follows a fractional Brownian motion.

The literature on long memory estimation by state space models is relatively barren. The existing literature takes two strands, namely estimation in the frequency domain and in the time domain. The two most popular estimation procedures includes the Geweke Porter-Hudak (1983) bandwidth estimation and the Local Whittle Gaussian model from Robinson (1995)². However, few empirical applications exist within state space models. Estimation of ARFIMA(p, d, q) processes by the Kalman filter was developed by Chan and Palma (1998) where they showed that by truncating the lag order of a long memory process, consistent estimates can be derived, provided that the series is stationary and invertible.

In the present study, time series of expected returns and dividend growth rate are derived assuming that expected returns has an ARFIMA(p, d, q) structure. The expected dividend growth rate is still assumed to follow an autoregressive process of order one. The results of KVB are reproduced in order to compare with ARFIMA model. Our model specifically studies the case of an ARFIMA($1, d, 0$)³. In the case of the AR(1), it is also shown that the model parameters may be derived by matching the moments of an ARMA(1, 1) to the theoretical model.

Expected returns, which is fractionally integrated, can be represented as infinite moving average or autoregressive process in the

²Many other models have been presented which make use of bandwidth or wavelets. An interesting survey may be found in Chan and Palma (2005).

³Many different specifications of the process may be attempted but it is computationally intensive. An ARFIMA (1,d,0) will take into account both the short and long range components.

time series model. One method to estimate ARFIMA(p, d, q) is to use the Chan and Palma (1998) model where the exact likelihood of the ARFIMA model is computed recursively by a Kalman Filter. In this chapter, we use a truncated autoregressive process to model the infinite representation of a long memory process. The corresponding state space model is an autoregressive process with noise. The truncated autoregressive process emulates the properties of an ARFIMA(p, d, q). A Monte Carlo experiment is reported where simulated ARFIMA($0, d, 0$) series are estimated by a truncated autoregressive process in the Kalman Filter. The log likelihood function of the Kalman filter is optimized to the current data set to yield the optimal parameters of the present value. As a by-product of the procedure, the expected returns and expected dividend (earnings) growth rate are used as predictors for realized returns and observed dividend (earnings) growth rate. The relationship between consumption and expected returns is also investigated. Finally, the expected returns and expected (earnings) growth rate is used in the present value reverting trading strategy.

A potential problem encountered when using dividend growth and the price-dividend ratio is that it may not be fully representative of payoffs in the presence of share repurchases. In this case, the Campbell-Shiller(1989) present value makes use of the price-earnings ratio and earnings growth in the present value relation. In a nutshell, I estimate the model for both the AR(1) and ARFIMA($1, d, 0$) using both dividend and earnings data for the time periods 1926-2008 and 1946-2008.

In the two step model, a sequential approach is used. In the first step, dividend growth is forecast in real-time assuming three least squares specifications, namely the simple mean model, the autoregressive model and a regression model made up of the present value

variables. One step ahead forecasts are produced from the regression models, which are in turn replaced in the present value formulation of Campbell and Shiller (1988). The expected returns are then solved for, and fitted with ARFIMA (p,d,q) processes. Tests of long memory and time variation are performed on the filtered series.

The remainder of the paper is organized as follows. Section 2 models the state space present value assuming AR(1) and ARFIMA(p,d,q) processes. Section 3 presents the 2 stage present value model. Section 4 reports and explains the results. Three applications of the filtered series are reported. Section 5 checks the robustness of the results and performs tests of persistence and time variation. Section 6 concludes.

5.2 Present Value assuming AR(1)

In this section, the AR(1) specification of expected growth and expected returns is reintroduced with the objective of deriving both the state space specification and the corresponding moments if the model was to be estimated by a simple OLS.

The key variables of the present value are the rate of return, price-dividend ratio and dividend growth:

$$r_{t+1} = \log\left(\frac{P_{t+1} + D_{t+1}}{P_t}\right). \quad (5.1)$$

$$PD_t = \frac{P_t}{D_t}. \quad (5.2)$$

$$\Delta d_{t+1} = \log\left(\frac{D_{t+1}}{D_t}\right). \quad (5.3)$$

The mean adjusted conditional expected returns and dividend growth rate are modelled as autoregressive processes of order one in (5.4) and (5.5) respectively :

$$\mu_{t+1} - \delta_0 = \delta_1(\mu_t - \delta_0) + \varepsilon_{t+1}^\mu, \quad (5.4)$$

$$g_{t+1} - \gamma_0 = \gamma_1(g_t - \gamma_0) + \varepsilon_{t+1}^g, \quad (5.5)$$

where $\mu_t = E_t(r_{t+1})$ and $g_t = E_t(\Delta d_{t+1})$. (5.4) and (5.5) show the mean deviation of expected returns and expected dividend growth where δ_0 and γ_0 characterize the unconditional mean of the expected returns and dividend growth respectively. δ_1 and γ_1 represent the autoregressive parameters. ε_{t+1}^μ and ε_{t+1}^g are shocks to the expected returns and the dividend growth rate processes. The shocks are normally distributed: $\varepsilon_{t+1}^\mu \sim N(0, \sigma_\mu^2)$ and $\varepsilon_{t+1}^g \sim N(0, \sigma_g^2)$.

The realized dividend growth rate is defined as the expected dividend growth rate and the unobserved shock ε_{t+1}^d , where by :

$$\Delta d_{t+1} = g_t + \varepsilon_{t+1}^d. \quad (5.6)$$

ε_{t+1}^d and g_t are assumed to be orthogonal to each other. $E(\varepsilon_{t+1}^d, g_t) = 0$.

The Campbell and Shiller (1988) log linearized return approximation (derived in appendix A.1) may be written as :

$$r_{t+1} \simeq \kappa + \rho pd_{t+1} + \Delta d_{t+1} - pd_t, \quad (5.7)$$

where $pd_t = \log(PD_t)$, $\overline{pd} = E(pd_t)$, $\kappa = \log(1 + \exp(\overline{pd})) - \rho \overline{pd}$ and $\rho = \frac{\exp(\overline{pd})}{1 + \exp(\overline{pd})}$.

By iterating (5.7) and using assumptions (5.4) and (5.5) , the

functional form of the process can be written as (5.8) after applying the expectations operator :

$$pd_t = A - B_1(\mu_t - \delta_0) + B_2(g_t - \gamma_0), \quad (5.8)$$

where $A = \frac{\kappa}{1-\rho} + \frac{\gamma_0 - \delta_0}{1-\rho}$, $B_1 = \frac{1}{1-\rho\delta_1}$ and $B_2 = \frac{1}{1-\rho\gamma_1}$.

5.2.1 Time Series Properties of Latent and Measured Variables

The latent parameters from the present value may be estimated using Ordinary Least Squares. In what follows, we shall derive some time series properties for expected dividend growth and show how the corresponding series may be derived by matching the moments of an ‘estimated’ model to that of a ‘theoretical’ model. The same conditions apply for expected returns since they both follow an AR(1). The demeaned form of the expected dividend growth may be written as (5.9)

$$\tilde{g}_{t+1} = \gamma_1 \tilde{g}_t + \varepsilon_{t+1}^g, \quad (5.9)$$

where $\tilde{g}_{t+1} = g_{t+1} - \gamma_0$.

Using the lag operator notation,

$$\begin{aligned} (1 - \gamma_1 L)\tilde{g}_{t+1} &= \varepsilon_{t+1}^g \\ \tilde{g}_{t+1} &= \frac{\varepsilon_{t+1}^g}{(1 - \gamma_1 L)} \\ \tilde{g}_t &= \frac{\varepsilon_t^g}{(1 - \gamma_1 L)} \end{aligned} \quad (5.10)$$

where L is the lag operator. Replacing (5.10) into (5.6) realized dividend growth may be written as (5.11):

$$\begin{aligned}\Delta d_t &= \gamma_0(1 - \gamma_1) + \gamma_1\Delta d_{t-1} + \varepsilon_{t-1}^g + \varepsilon_t^d \\ &\quad - \gamma_1\varepsilon_{t-1}^d\end{aligned}\tag{5.11}$$

$$= \bar{\gamma} + \gamma_1\Delta d_{t-1} + \eta_t^{dg},\tag{5.12}$$

where $\bar{\gamma} = \gamma_0(1 - \gamma_1)$ and $\eta_t^{dg} = \varepsilon_{t-1}^g + \varepsilon_t^d - \gamma_1\varepsilon_{t-1}^d$. (5.11) has an ARMA (1, 1) structure if it is to be estimated. However, the structural parameters (γ_0 , γ_1 , σ_d^2 and σ_g^2) cannot be directly inferred from this model. To compute these values, the theoretical moments may need to be matched to the empirical counterparts. The autocovariance, variance and autocorrelation of η_t^{dg} are specified as follows:

Autocovariance at one lag:

$$\gamma_0(1) = E(\eta_t^{dg}\eta_{t-1}^{dg}) = -\gamma_1\sigma_d^2.$$

Variance:

$$\text{var}(\eta_t^{dg}) = \sigma_g^2 + (1 + \gamma_1^2)\sigma_d^2.\tag{5.13}$$

Autocorrelation at one lag:

$$\rho_{\eta_t^{dg}}(1) = \frac{-\gamma_1\sigma_d^2}{\sigma_g^2 + (1 + \gamma_1^2)\sigma_d^2}.\tag{5.14}$$

Moreover the variance of the observed dividend growth is:

$$\text{var}(\Delta d_t) = \frac{\sigma_g^2 + (1 - \gamma_1^2)\sigma_d^2}{(1 - \gamma_1^2)\sigma_d^2}.\tag{5.15}$$

(5.11) can be described as the **theoretical** model of observed dividend growth. A conventional way for estimating the structural parameters is to estimate an ARMA (1, 1). η_t is defined as a moving average:

$$\eta_t = v_t = u_t - \theta u_{t-1},$$

where $u_t \sim N(0, \sigma_u^2)$. In this case, the model for **estimation** of dividend growth is given by:

$$\Delta d_t = \gamma_0(1 - \gamma_1) + \gamma_1 \Delta d_{t-1} + u_t - \theta u_{t-1}. \quad (5.16)$$

This process has the following variance, autocovariance and autocorrelation:

$$\text{var}(v_t) = (1 + \theta^2)\sigma_u^2 \quad (5.17)$$

$$E(v_{t-1}v_t) = -\theta\sigma_u^2$$

$$\rho_v(1) = \frac{-\theta}{(1 + \theta^2)} \quad (5.18)$$

Since the autocorrelation from the theoretical (5.15) and estimated (5.18) model are equal, the matched moments imply :

$$\frac{-\gamma_1\sigma_d^2}{\sigma_g^2 + (1 + \gamma_1^2)\sigma_d^2} = \frac{-\theta}{(1 + \theta^2)}.$$

The signal to noise ratio $R = \frac{\sigma_g^2}{\sigma_d^2}$, is therefore equal to :

$$R = \gamma_1 \left(\frac{1}{\theta} + \theta \right) - (\gamma_1^2 + 1).$$

The variance of expected dividend growth can be written in terms of the estimated parameters with the observed dividend growth:

$$\sigma_g^2 = [\gamma_1 \left(\frac{1}{\theta} + \theta \right) - \gamma_1^2 - 1]\sigma_d^2. \quad (5.19)$$

Replacing (5.19) into (5.13), and equating it to the estimated variance of v_t (5.17) will yield:

$$[\gamma_1 \left(\frac{1}{\theta} + \theta \right) - \gamma_1^2 - 1] \sigma_d^2 + (1 + \gamma_1^2) \sigma_u^2 = (1 + \theta^2) \sigma_u^2.$$

This implies that

$$\sigma_d^2 = \frac{\theta \sigma_u^2}{\gamma_1} \quad (5.20)$$

(5.20) implies that σ_d^2 can be estimated directly, since θ and γ_1 are known (estimated from equation 5.16). σ_g^2 can therefore be computed as follows:

$$\sigma_g^2 = [\gamma_1 \left(\frac{1}{\theta} + \theta \right) - \gamma_1^2 - 1] \frac{\theta \sigma_u^2}{\gamma_1}. \quad (5.21)$$

Both σ_d^2 and σ_g^2 can therefore be computed. γ_0 may be computed by matching the estimated intercept term.

State Space Model

The state space model makes use of a transition equation and a measurement equation. The Kalman Filter illustrates the dynamics of the series of μ_t and g_t . The two transition equations are given by:

$$\tilde{g}_{t+1} = \gamma_1 \tilde{g}_t + \varepsilon_{t+1}^g, \quad (5.22)$$

$$\tilde{\mu}_{t+1} = \delta_1 \tilde{\mu}_t + \varepsilon_{t+1}^\mu. \quad (5.23)$$

The two measurement equations are given by :

$$\Delta d_{t+1} = \gamma_0 + \tilde{g}_t + \varepsilon_{t+1}^d, \quad (5.24)$$

$$pd_t = A - B_1 \tilde{\mu}_t + B_2 \tilde{g}_t. \quad (5.25)$$

(5.24) is the equation linking the observed dividend growth rate to the state variable of expected dividend growth. (5.25) is the

present value equation linking price-dividend to expected returns and expected dividend growth rate. It is a generalization of equation 5.7, where $A = \frac{\kappa}{1-\rho} + \frac{\gamma_0 - \delta_0}{1-\rho}$, $B_1 = \frac{1}{1-\rho\delta_1}$ and $B_2 = \frac{1}{1-\rho\gamma_1}$.

Equation (5.25) can be rearranged into (5.28) such that there are only two measurement equations and only one state equation.

$$\tilde{g}_{t+1} = \gamma_1 \tilde{g}_t + \varepsilon_{t+1}^g, \quad (5.26)$$

$$\Delta d_{t+1} = \gamma_0 + \tilde{g}_t + \varepsilon_{t+1}^d, \quad (5.27)$$

$$pd_{t+1} = (1 - \delta_1)A + B_2(\gamma_1 - \delta_1)\tilde{g}_t + \delta_1 pd_t - B_1 \varepsilon_{t+1}^\mu + B_2 \varepsilon_{t+1}^g. \quad (5.28)$$

(5.26) defines the transition (state) equation. The measurement equations are given by (5.27) and (5.28). Since all the equations are linear, we can implement the Kalman Filter and obtain the likelihood which is maximized over the following vector of parameters. The likelihood is optimized using the MaxBFGS procedure in Ox.

$$\Theta = (\gamma_0, \delta_0, \gamma_1, \delta_1, \sigma_g, \sigma_\mu, \sigma_D, \rho_{g\mu}, \rho_{gd}, \rho_{\mu d}) \quad (5.29)$$

5.3 Present Value assuming an ARFIMA(p, d, q)

In this section, the ARFIMA(p, d, q) model is introduced. The ARFIMA(p, d, q) is shown to follow an autoregressive process of infinite order. The autoregressive process is truncated of order m . The structural model is then estimated by the Kalman Filter.

We shall briefly introduce some estimation models which are popular in the economics and finance literature. Interesting studies

in the field of parametric estimation include Granger and Joyeux (1980), Fox and Taqqu (1986), Dahlhaus (1989) Sowell (1992) and Beran (1994a). In the semiparametric case, good contributions include Geweke and Porter-Hudak (1983), Robinson (1995a) and Moulines and Soulier(2001). Geweke and Porter-Hudak (1983) is perhaps one of the most important estimators of long memory where it is based on the behaviour of spectral density of fractional processes close to zero. They developed an estimator of d based on the simple regression on the logarithm of periodogram points. Sowell (1992) and Lobato and Robinson (1996) use the maximum likelihood procedure to simultaneously estimate $p, d,$ and q .

The Kalman Filter estimation of long memory model has not yet been popularized in economics but has been considered in mathematical and physical modeling field. We shall first explain the principles guiding the framework before applying the state space representation to the present value. An ARFIMA process may be transformed into an autoregression and moving average of infinite order. For the Kalman Filter estimation, the autoregressive process is truncated of order m . The autoregressive lags in the AR process must be of an order high enough in order to capture the theoretical autocorrelation structure of the original model. However, the corresponding trade-off is the computational power involved in estimating a high order autoregressive process. Examples of approximating ARFIMA models by ‘long’ AR include Ray (1993), Ray and Crato (1996), Chan and Palma (1998), Poskitt (2006), Lahiani and Scaillet (2008).

An ARFIMA(p, d, q) process, x_t for $t = 1 \dots T$ may be written as:

$$\varphi(L)(1 - L)^d x_t = \theta(L)\eta_t, \quad (5.30)$$

where η_t is white noise: $(\eta_t \sim N(0, \sigma_\eta^2))$.⁴

$\varphi(L)$ and $\theta(L)$ are respectively equal to $(1 - \varphi_1 L - \varphi_2 L^2 - \dots - \varphi_p L^p)$ and $(1 - \theta_1 L - \theta_2 L^2 - \dots - \theta_q L^q)$. $\varphi(L)$ and $\theta(L)$ are also assumed to have no common factor.

The expansion of $(1 - L)^d$ may be expressed as follows:

$$\begin{aligned} (1 - L)^d &= \sum_{j=0}^{\infty} \lambda_j L^j & (5.31) \\ &= \sum_{j=0}^{\infty} \binom{d}{j} (-L)^j \\ &= \lambda(L), \end{aligned}$$

where $\lambda_j = \frac{\Gamma(j-d)}{\Gamma(-d)\Gamma(j+1)}$.

5.3.1 Autoregressive Approximation to the ARFIMA(p, d, q)

Some properties of parameters are outlined. When $d < |\frac{1}{2}|$, the roots of $\varphi(L)$ and $\theta(L)$ lie outside the unit circle. When $0 < d < 0.5$, x_t exhibits positive dependence between distant observations. $d = 0$ implies that the process has short memory, and is simply an ARMA(p, q). When $-0.5 < d < 0$, the series exhibits negative dependence. The case of ARFIMA($0, d, 0$) is known as a fractional noise process. (5.30) may be rewritten in the following form.

$$\eta_t = \frac{\theta(L)}{\varphi(L)} (1 - L)^d x_t. \quad (5.32)$$

Replacing (5.31) in equation (5.32):

$$\eta_t = \frac{\theta(L)}{\varphi(L)} \sum_{j=0}^{\infty} \lambda_j L^j x_t,$$

where $\lambda_0 = 1, \lambda_1 = -d, \lambda_2 = \frac{d(1-d)}{2}, \lambda_j = \lambda_{j-1}(j-1-d)/j$

⁴ x_t may denote either expected returns or expected dividend growth rate.

$$\eta_t = \frac{\sum_{i=1}^p \varphi_i L^i \sum_{j=0}^{\infty} \lambda_j L^j}{\sum_{i=1}^q \theta_i L^i} x_t.$$

The process may be written as:

$$\left(1 - \sum_{j=1}^{\infty} \pi_j L^j\right) x_t = \eta_t, \quad (5.33)$$

where $\pi_i = \lambda_i - \sum_{j=1}^q \theta_j \pi_{i-j} + \sum_{j=1}^p \varphi_j \lambda_{i-j}$. (5.33) is an autoregressive process of infinite order.

Truncated Autoregressive Specification

Estimating model (5.33) may be computationally demanding, and theoretically impossible in a finite sample. For estimation purposes, the lag order of the infinite autoregression is truncated. The truncation level shall be denoted by m . In the context of measuring the persistence of a series, a consistent estimate of the model parameters requires that the truncated AR(m) specification capture the autocorrelation structure of the ARFIMA(p, d, q). m should be high ideally. However, the cost of adopting a high m is the computational expense involved. The AR(m) is given by equation (5.34):

$$x_t = \eta_t + \pi_1 x_{t-1} + \pi_2 x_{t-2} + \cdots + \pi_m x_{t-m}, \quad (5.34)$$

where $m \leq t \leq T$. π_j for long memory models is given by the iteration $\pi_{j-1}(j-d-1)/j$, with $\pi_0 = 1$.

Long autoregressive processes have been used to approximate ARFIMA in the forecasting literature. Although the truncated AR(m) may not capture long memory properties in the strict sense, estimation of AR(m) essentially attempts to model the hyperbolic

decay with the sum of exponential decays in the autocorrelation structure. In other words, the autocorrelation for the truncated $AR(m)$ is equal to that of the $ARFIMA(p, d, q)$, such that the tail of the $AR(\infty)$ does not matter. As m increases, we should expect the fit of the $ARFIMA(p, d, q)$ to improve as well. Fitting the truncated $AR(m)$ for $m = 50$ from a sample of $T = 348$, for a computed $d = 0.417$ is shown by Geweke and Porter-Hudak to provide forecasts which emulate those of fractional noise models.

State space framework : *AR-plus-noise Model*

Using the truncated approach, (5.32) may be put in state space form. (5.34) may be written as :

$$\pi(L)x_t = \eta_t, \tag{5.35}$$

where $\pi(L) = 1 - \pi_1 L - \dots - \pi_m L^m$.

The state space system requires a state equation and a transition equation. The state equation is the truncated autoregressive process (5.35) and the measurement equation is a one to one relationship between the observed variable (y_t) and the state variable (x_t) as in (5.36).

$$\begin{aligned} \pi(L)x_t &= \eta_t \\ y_t &= x_t + \varepsilon_t^y, \end{aligned} \tag{5.36}$$

where $\varepsilon_t^y \sim NID(0, \sigma_{\varepsilon^y}^2)$ and both ε_t^y and η_t are uncorrelated. When a matrix structure is imposed, the model is given by (5.37) and (5.38).

$$x_t = Fx_{t-1} + \eta_t, \quad (5.37)$$

$$y_t = Gx_t + \varepsilon_t^y, \quad (5.38)$$

where the state vector is given by

$$x_t = \begin{bmatrix} x_t & x_{t-1} & \dots & x_{t-m+2} & x_{t-m+1} \end{bmatrix}'. \quad (5.39)$$

The elements of matrices F and G are given as follows:

$$F = \begin{bmatrix} \pi_1 & \pi_2 & \dots & \pi_{m-1} & \pi_m \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & & & & \vdots \\ 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix} \quad G = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \end{bmatrix}.$$

The state matrix F is represented such that (5.37) holds. The ones in the lower triangular diagonal ensure trivial identities such that the left hand side is equal to the right hand side. The reliability of the truncated *AR-plus-noise Model* is investigated via a small Monte Carlo experiment with 100 runs for truncated $m = 5, 10, 20$ and 40 for $d = 0.05, 0.2$ and 0.4 .⁵ The different values of T were 100 and 200. The results from the Monte Carlo experiment are illustrated in table A.17 in the appendix. The bias is not excessively high for a small sample of $T = 100$ and a truncation lag of $m = 5$. There is convergence towards the true d for larger m .

State Space: *Present Value*

In this section, we show the state space representation for the present value. The state equations for expected dividend growth and the

⁵Monte Carlo experiments with a long AR is computationally very expensive, explaining the small number of runs.

expected returns are given as follows:

$$x_{\mu,t+1} = \delta_0 + F_{\mu}x_{\mu,t} + \eta_{\mu,t+1}, \quad (5.40)$$

$$x_{g,t+1} = \gamma_0 + F_g x_{g,t} + \eta_{g,t+1}, \quad (5.41)$$

where $x_{\mu,t+1}$ and $x_{g,t+1}$ are the state variables for expected returns and expected dividend growth respectively and have structure as 5.39. The structure (5.41) also accounts for the possibility of long memory in expected dividend growth. δ_0 and γ_0 are scalars defining the unconditional means of expected returns and dividend growth respectively. Similar to the AR(1) case, $\eta_{\mu,t}$ and $\eta_{g,t}$ may be correlated over time.

Measurement equation

The log linearized return is:

$$r_{t+1} \simeq \kappa + \rho pd_{t+1} + \Delta d_{t+1} - pd_t, \quad (5.42)$$

with $\bar{pd} = E(pd_t)$, $\kappa = \log(1 + \exp(\bar{pd})) - \rho pd$, and $\rho = \frac{\exp(\bar{pd})}{1 + \exp(\bar{pd})}$. Assuming the transversality condition and applying conditional expectations, the present value approximation may be rewritten as (5.43):

$$pd_t \simeq \frac{\kappa}{1 - \rho} + b(x_{g,t} - x_{\mu,t}), \quad (5.43)$$

where $b = \frac{1}{1 - \rho} \cdot \rho$ and κ are as defined as in the autoregressive section.

From (5.38), the observed dividend growth is equal to :

$$\Delta d_{t+1} = \gamma_0 + G' x_{g,t} + \eta_{t+1}^d.$$

Hence the state space model is equal to the following measurement equations:

$$\begin{aligned}\Delta d_{t+1} &= \gamma_0 + G'x_{g,t} + \eta_{d,t+1}, \\ pd_{t+1} &\simeq A + b[F_g x_{g,t} + \eta_{g,t+1} - F_\mu x_{\mu,t} - \eta_{\mu,t+1}].\end{aligned}$$

where $A = \frac{\kappa + \gamma_0 - \delta_0}{1 - \rho}$. The state space may therefore be summarized as:

$$\mathbf{X}_{t+1} = \mathbf{A}_0 + \mathbf{A}_1 \mathbf{X}_t + \boldsymbol{\eta}_{1,t+1}, \quad (5.44)$$

$$\mathbf{Y}_{t+1} = \mathbf{A}_2 + \mathbf{A}_3 \mathbf{X}_t + \boldsymbol{\eta}_{2,t+1}, \quad (5.45)$$

where:

$$\begin{aligned}\mathbf{X}_t &= \begin{bmatrix} x_{\mu,t} & x_{g,t} \end{bmatrix}', \\ \mathbf{Y}_t &= \begin{bmatrix} \Delta d_t & pd_t \end{bmatrix}', \\ \boldsymbol{\eta}_{1,t} &= \begin{bmatrix} \eta_{\mu,t} & \eta_{g,t} \end{bmatrix}', \\ \boldsymbol{\eta}_{2,t} &= \begin{bmatrix} \eta_{d,t} & 0 \end{bmatrix}', \\ \mathbf{A}_0 &= \begin{bmatrix} \delta_0 & \gamma_0 \end{bmatrix}', \\ \mathbf{A}_1 &= \begin{bmatrix} F_\mu & 0 \\ 0 & F_g \end{bmatrix}, \\ \mathbf{A}_2 &= \begin{bmatrix} \delta_0 & 0 \end{bmatrix}', \\ \mathbf{A}_3 &= \begin{bmatrix} 0 & G' \\ -bF_\mu & bF_g \end{bmatrix}.\end{aligned}$$

The standard Kalman Filter is applied to the system of equation (5.44) and (5.45). The vector to be optimized is given by:

$$\Theta = (\gamma_0, \delta_0, \gamma_1, \delta_1, \sigma_g, \sigma_\mu, \sigma_d, \rho_{g\mu}, \rho_{gd}, \rho_{\mu d}, d_\mu, d_g). \quad (5.46)$$

5.4 Two Step Present Value Model

In this section, I explain a very simple procedure to derive expected returns in real time. The assumption borrowed in this model is that agents use econometric forecasts to form their expectations. I assume three simple regression schemes to forecast dividend one step ahead⁶. In the second step, a series of expected returns are derived from the present value model in a time varying framework.

Consider the Campbell and Shiller (1988) log linearized present value approximation from (5.7):

$$r_{t+1} \simeq \kappa + \rho pd_{t+1} + \Delta d_{t+1} - pd_t.$$

By iterating the process infinite steps forward,

$$pd_t = \frac{\kappa}{1-\rho} + \rho^\infty pd_\infty + \sum_{i=1}^{\infty} \rho^{i-1} (\Delta d_{t+i} - r_{t+i}).$$

Assuming the no bubble condition, $\lim \rho^{t+n} pd_{t+n} = 0$, as n tends to infinity:

$$pd_t = \frac{\kappa}{1-\rho} + \sum_{i=1}^{\infty} \rho^{i-1} (\Delta d_{t+i} - r_{t+i}).$$

Putting the conditional expectations operator on both sides of the model, the equation can be rewritten as:

$$pd_t = E_t\left(\frac{\kappa}{1-\rho}\right) + E_t\left(\sum_{i=1}^{\infty} \rho^{i-1} (\Delta d_{t+i} - r_{t+i})\right) \quad (5.47)$$

$$pd_t = E_t\left(\frac{\kappa}{1-\rho}\right) + C_1 [\bar{g}_t - \bar{\mu}_t]. \quad (5.48)$$

where $C_1 = \frac{1}{1-\rho}$.

⁶ More complicated models, which can accommodate nonlinearities may yield spurious forecasts given the present sample size.

In the literature, κ and ρ are assumed to be constant such that $E_t(\frac{\kappa}{1-\rho}) = \frac{\kappa}{1-\rho}$. In this paper, this assumption is refined to accommodate time variation, such that κ, ρ and C_1 are all time varying (κ_t, ρ_t and C_{1t}). I show in the next section that κ, ρ and C_1 should be time varying as they tend to be cyclical in nature.

The two step procedure is simple. In the first step, \bar{g}_t is estimated by three simple regression models. This involves forecasting future realized dividend growth using the available information set. To that end, all the estimated models are forecast on a recursive window basis. The three regression models used to generate \bar{g}_t are as follows:

Model 1 is a simple constant mean regression model. The agent observes the dividend growth in the past periods and decides to forecast dividends on the prevailing mean α . The dividend growth forecast for period $t + 1$, but produced at time t is thus given by $t^{-1} \sum_{j=1}^t \Delta d_j$ under the recursive window.⁷

$$\bar{g}_t = E(\Delta d_{t+1} | \mathcal{F}_t) = \alpha + \varepsilon_{t+1},$$

where ε_{t+1} is the forecast error and \mathcal{F}_t is the information set at time t .

Model 2 assumes that dividend growth is forecast from an AR(1) process. Higher orders of the autoregression may be included but given the frequency and the unpredictable nature of dividend growth, an AR(1) seems reasonable.

$$\bar{g}_t = E(\Delta d_{t+1} | \mathcal{F}_t) = \alpha + \beta_1 \Delta d_{t-1} + \varepsilon_{t+1}.$$

Model 3 is a factor model which includes lagged regressors from the Campbell-Shiller equation. In this setting, both the price-dividend

⁷An interesting feature of this model is that under the recursive window, dividend growth is not predictable as it has a constant mean. This may be contrasted with rolling windows which posits that there is a slowly changing mean.

ratio and the realized returns enter the equation

$$\bar{g}_t = E(\Delta d_{t+1} | \mathcal{F}_t) = \alpha + \beta_1 p d_t + \beta_2 r_t + \varepsilon_{t+1}.$$

In all the three models, the error term ε_{t+1} is treated as white noise. In the second step, $\bar{\mu}_t$ is generated from the present value approximation (5.49)

$$\bar{\mu}_t = \frac{1}{C_{1t}} \left[p d_t - \frac{\kappa_t}{1 - \rho_t} - C_{1t} \bar{g}_t \right]. \quad (5.49)$$

5.4.1 Time Variation in κ , ρ and C_1

The Campbell and Shiller approximation may be interpreted as a difference equation which relates the long run coefficients κ , ρ and the short run variables \bar{g}_t and $\bar{\mu}_t$. However, intuitively, the long run coefficients are made up of the short run coefficients and may be rewritten as (5.50):

$$\frac{\kappa}{1 - \rho} \approx p d_t - C_1 [\bar{g}_t - \bar{\mu}_t]. \quad (5.50)$$

(5.50) simply shows that the log linearization parameter may be interpreted as a function of t. This may be interpreted in contemporary macroeconomics as a ‘moving equilibrium’ over time. To account for this possibility we shall allow κ and ρ to depend on time. The graphical plots of κ show that the latter is not constant over time. κ for dividend growth and earnings growth is plotted in A.15 and A.16 in the appendix.

5.5 Results

5.5.1 Data

The optimization model uses earnings, dividends, price and consumption data from 1926-2008. The returns series is derived by taking the log difference of the price series. Earnings and Dividend growth are measured as the logarithmic difference between respective payoffs from time $t - 1$ to t .

5.5.2 Optimization of State Space Models

In this section, I report the results of the AR(1) and ARFIMA(1,d,0) defined in the previous sections for the sample period 1926-2008 and 1946-2008 for both the dividends and earnings.

Table 5.1:

Estimation of AR(1) and ARFIMA(1,d,0) Model for Dividend data.

The parameters optimized are from the previously defined parameter sets 5.29 (AR(1)) and 5.46 (ARFIMA(1,d,0)) over the two periods. The ‘d’ parameter for the autoregressive expected dividend growth process is set to zero.

Price-dividend Ratio/Dividend Growth								
1926-2008					1946-2008			
Parameters	AR(1)		ARFIMA(1,d,0)		AR(1)		ARFIMA(1,d,0)	
	PARAM	SE	PARAM	SE	PARAM	SE	PARAM	SE
γ_0	0.021	0.014	0.003	0.081	0.019	0.012	0.018	0.001
δ_0	0.055	0.019	0.029	0.074	0.046	0.020	0.05	0.001
γ_1	0.11	0.119	0.165	0.210	0.395	0.203	0.326	0.108
δ_1	0.921	0.050	0.158	0.031	0.929	0.049	0.118	0.033
d_1	–	–	0.475	0.050	–	–	0.457	0.606
σ_g	0.052	0.016	0.109	0.001	0.05	0.014	0.048	0.193
σ_μ	0.02	0.010	0.044	0.001	0.015	0.009	0.044	0.030
σ_d	0.092	0.055	0.007	0.001	0.014	0.040	0.007	0.043
$\rho_{g\mu}$	0.576	0.088	0.212	0.001	0.62	0.126	0.345	0.494
$\rho_{\mu d}$	-0.046	0.001	-0.166	0.001	-0.055	0.685	-0.002	0.250
Log-Likelihood	-103.79		-110.45		-126.83		-130.07	

According to the log likelihood value, the ARFIMA(1, d , 0) tends to perform better than the AR(1) except for the 1946-2008 dividend sample. Generally, the ARFIMA model is superior when using earnings. The result is fairly simple to understand within the present value framework. The price-earnings ratio and the price-dividend ratio share a common level of persistence (i.e. An autoregression on the price-dividend ratio and price-earnings ratio produces nearly the same autoregressive parameter). However the observed dividend growth is much more persistent than earnings growth. According to approximation (5.25), the expected returns in earnings equations should have a higher degree of persistence which is adequately represented by an ARFIMA(1,d,0).

The optimized results have the same near unit root properties

Table 5.2:

Estimation of AR(1) and ARFIMA(1,d,0) Model for Earnings data.

In this table, the measurement variables, price-dividend ratio and dividend growth, are replaced by the price-earnings ratio and the earnings growth respectively.

Price-earnings Ratio/ Earnings Growth								
1926-2008					1946-2008			
Parameters	AR(1)		ARFIMA(1,d,0)		AR(1)		ARFIMA(1,d,0)	
	PARAM	SE	PARAM	SE	PARAM	SE	PARAM	SE
γ_0	0.023	0.007	0.010	0.038	0.006	0.035	0.02	0.042
δ_0	0.089	0.044	0.079	0.054	0.048	0.068	0.05	0.052
γ_1	0.082	0.001	0.089	0.049	0.101	0.128	0.457	0.001
δ_1	0.878	0.001	0.101	0.042	0.927	0.048	0.382	0.001
d_1	—	—	0.363	0.045	—	—	0.499	0.001
σ_g	0.197	0.001	0.228	0.014	0.218	0.023	0.031	0.001
σ_μ	0.057	0.001	0.101	0.012	0.043	0.017	0.024	0.001
σ_e	0.116	0.001	0.184	0.020	0.016	0.110	0.045	0.001
$\rho_{g\mu}$	0.87	0.001	0.829	0.065	0.929	0.043	0.237	0.001
$\rho_{\mu e}$	0.126	0.001	-0.198	0.069	-0.364	0.273	0.284	0.001
Log-Likelihood	-35.27		-113.03		-39.99		-123.61	

as in KVB for the autoregressive processes. The unit root is found in both samples with the price-dividend ratio and also when the price-earnings ratio is used. Interestingly, the memory component (d) is high in almost all four models. The short range component (autoregressive part) of the ARFIMA tends to be lower. Dividend growth tends to have a similar short run parameter over both periods. The parameters for the dividend growth equation changes only marginally. The sample 1946-2008 is associated with an equal improvement in the autoregressive parameter for both expected dividend growth and expected earnings growth.

For both samples, the variation of the expected and realized dividend and earnings growth rate tend to be higher than that of expected returns. The expected earnings growth appears to vary much

more than the dividend growth for both samples and both models. The correlation between expected earnings growth and expected returns tends to be stronger than that between dividend growth and expected returns. Some of these findings can be confirmed from tables A.22, A.23, A.24 and A.25 in the appendix. The mean values of the expected returns and expected dividend growth rate are close to the observed. The 1926-2008 earnings growth rate, however are exceptionally high. Interestingly, tests of stationarity $I(0)$ and non-stationarity $I(1)$ show that the AR(1) models tend to be closer to being non-stationary, unlike ARFIMA models. The Robinson-Lobato p-values show that upon the adoption of the ARFIMA model, the p-values signalling rejection of fractional alternatives, tend to be lower than the AR(1) model, hence putting the case forward for a fractional process.

Robustness Checks

Robustness over time Robustness checks were performed by comparing the estimated parameters over the two samples. The estimated parameters of the ARFIMA(1, d , 0) seem to be significantly different over the two sample. Although dividend growth appears more stable than earnings, significant differences among the parameters γ_0 , γ_1 and $\rho_{\mu d}$ are witnessed. In terms of the earnings growth, most of the parameters appear to be unstable.

This may be due to two different reasons. Firstly, the model is based on a small sample where many parameters have to be estimated. The model makes use of only a few data points (84 years) for the estimation of 10 parameters. Secondly, different regimes within the two samples may lead to substantial variation of the ARFIMA process. Based on the latent nature of the expected returns, a method to check for robustness is to see whether expected returns across the different time periods exhibit high correlation. The pairwise correlation between the various specifications of the returns series are reported in tables A.26 and A.27 in the appendix. In the case of the AR(1), most of the pairwise correlations are above 0.8. Both the price-dividend ratio and the price-earnings ratio exhibit a 0.99 correlation over the two different sample sizes.

The ARFIMA model tends to exhibit low correlation over time. However, earnings and dividend measures tend to demonstrate the same level of correlation in a specific sample. It is clear that the 1929-36 depression may have had led to higher expected returns. A Monte Carlo experiment was performed by using the same sample size (83 observations) for a two state Markov Regime Switching model where two states were considered: a high and a low level of expected returns. The Monte Carlo results, reported in table A.28 in the appendix, show that the ARFIMA parameters tend to exhibit a

higher variance than the simple AR process. Moreover the estimate of the intercept term is very unstable.

Univariate Models In this section, the expected returns and expected dividend (earnings) growth rates are modeled according to their initial exogenously determined econometric specifications. I present the results for the series for the individual samples in tables A.29, A.31, A.30 and A.32 in the appendix. The state space required looking at the dynamics of both dividend growth and returns at the same time. The corresponding univariate models ignores such behaviour. In other words, it sets the assumption $E_t(\mu_t g_t) = 0$.

The results show that the ARFIMA tends to perform worse than the AR(1) in the case of the expected returns (tables A.29 and A.30). The memory parameters tend to be unstable for both specifications of the autoregressive process. The ARFIMA tends to generally display a lower R-squared. The best ARFIMA specification is the earnings data for the sample 1946-2008. To a lesser extent, the ARFIMA removes dependence in the residual and reduces the ARCH effects. Interestingly, the univariate models for the dividend and earnings growth show more promise for the ARFIMA model. The linear fit of the model is greatly improved and the models are free from any serial correlation and conditional heteroscedasticity. The good fit of the model is perfectly clear within the present value approach. A lower fit in either the dividend growth rate or expected returns would improve the fit of the other variable, so that the persistence in the price-dividend ratio is restored.

Tests of Persistence and Time variation

In the following section, we detail some of the tests that were performed on the new series. First we test whether expected returns is a long memory process (i.e. $d > 0$). In terms of the definition of ‘persistence’, this is the strong form case because it takes into account the long range dependent case only. Secondly, we test the weaker form of persistence, where the autoregressive coefficients and the ‘d’ parameter are set to zero. For the dividend growth series, it involves looking at the autoregressive parameter. We also test for time variation in expected returns and expected dividend (earnings) growth rate. In this case, under the null hypothesis, the autoregressive parameter and the standard error of the transition equation shock are equal to zero.

The tests involve computing the likelihood ratio under alternative (L_1) and the null (L_0). The likelihood ratio test is computed as follows:

$$LR = 2(L_1 - L_0)$$

The likelihood ratio is distributed as $\chi^2(k)$ where k represents the number of restrictions.

The tests are performed exclusively on the ARFIMA(1,d,0) specification and are reported in table 5.3.

The log likelihood value will vary across the different samples and between earnings and dividend growth. However the results clearly demonstrate that the null hypothesis is being rejected in all cases. Expected returns do appear to exhibit long memory. There appears to be persistence in both the filtered returns and filtered dividend growth rate series. However the former exhibits a higher statistic in the region of the rejection of the null hypothesis, implying that there is a higher degree of persistence. The expected returns

Table 5.3:

Tests of Time Variation and Persistence.

The table shows the bootstrapped χ^2 for the different hypotheses. In the first case, the null hypothesis is $d = 0$, which implies that there is no memory at all. In the second case, both long and short range components are set to zero, under the null hypothesis. In the case of the time variation test, the corresponding standard errors are also set to zero.

Null Hypotheses	1926-2008 PD	1946-2008 PD	1926-2008 PE	1946-2008 PE
Memory $d = 0$	3368	2155	623	2690
Persistence Tests:				
$H_0 : \delta_1 = d = 0$	6478	3465	579	4153
$H_0 : \gamma_1 = 0$	22	12	14	458
Time Variation tests:				
$H_0 : \delta_1 = d = \sigma_\mu = 0$	17507	61797	317	4919
$H_0 : \gamma_1 = \sigma_g = 0$	1070	1806	130	684

series shows that there is enough joint evidence of a non zero d and the autoregressive parameter δ_1 . Tests for time variation show that both the expected returns and dividend growth rate tend to vary over time. However a naive comparison of the test statistic shows that expected returns exhibit more variation over time.

5.5.3 Applications

In this section, we provide three applications for the filtered returns and dividend (earnings) series. In the first application, we test for in-sample predictability. In this setting the filtered series are regressed on the realized values and the accuracy is measured by the R-squared. In the second application, we look at the effect of expected returns (as a proxy for discount rates) on consumption and consumption growth. Lastly, we test whether a trading strategy may be implemented by looking whether prices revert to their present value. We use the series for expected returns and expected

dividend (earnings) growth to construct the present value.

In-sample Predictability

The in-sample predictability of the realized series by the filtered series is reported in the tables 5.4 and 5.5. The following forecasting equations were run in the case of realized values of returns:

$$\begin{aligned}r_t &= \phi_o + \phi_1\mu_{t-1}^{AR} + \varepsilon_t \\r_t &= \phi_o + \phi_1\mu_{t-1}^{ARFIMA} + \varepsilon_t \\r_t &= \phi_o + \phi_1pd_{t-1} + \varepsilon_t\end{aligned}$$

For the dividend growth the following functional models were assumed:

$$\begin{aligned}\Delta d_t &= \phi_o + \phi_1g_{t-1}^{AR} + \varepsilon_t \\ \Delta d_t &= \phi_o + \phi_1g_{t-1}^{ARFIMA} + \varepsilon_t \\ \Delta d_t &= \phi_o + \phi_1pd_{t-1} + \varepsilon_t\end{aligned}$$

Returns are better forecast by price-dividend and price-earnings ratio in the samples 1926-2008 PD, 1926-2008 PE and 1946-2008 PE. The expected returns is marginally weaker than the ratios. The autoregressive process for the 1946-2008 PD sample appears to have good predictability. There is no apparent predictability for the expected dividend growth rate. The ARFIMA model performs relatively well for the sample 1946-2008 for the price-earnings ratio. However, expected dividend growth turns out to be a good predictor of the realized dividend growth.

Table 5.4:

Goodness of Fit for Returns Equation.

The figures in the table show the R-squared from using each regressor: the lagged filtered expected returns from the AR(1) and ARFIMA(1,d,0) and the price-dividend and price-earnings ratio. The dependent variable is the realized returns.

	AR(1)	ARFIMA(1,d,0)	pd_{t-1}/pe_{t-1}
1926-2008 PD	0.049	0.015	0.050
1946-2008 PD	0.101	0.088	0.066
1926-2008 PE	0.042	0.002	0.055
1946-2008 PE	0.072	0.079	0.090

Table 5.5:

Goodness of Fit for Dividend (Earnings) Growth Equation.

The figures in the table show the R-squared from each regressor. The regressors are the lagged filtered dividend or earnings growth and the price-dividend (earnings) ratio. In this case, the dependent variable is the realized dividend growth or earnings growth rate.

	AR(1)	ARFIMA(1,d,0)	pd_{t-1}/pe_{t-1}
1926-2008 PD	0.020	0.01	0.002
1946-2008 PD	0.015	0.03	0.008
1926-2008 PE	0.014	0.01	0.001
1946-2008 PE	0.027	0.08	0.002

Consumption and Expected Returns

The second application of both series is to see the reaction of consumption growth to a shock in expected returns. There is a wide theoretical literature linking the time series properties of consumption and discount rates (See Campbell (2003) and Cochrane (2010) for an overview). Consumption and discount rates are counter cyclical to each other. When discount rates (expected returns) are low, consumption is high. To test this relationship, a simple regression regressing expected returns on logarithm of consumption.

$$\ln C_t = \alpha + \beta\mu_t + v_t. \quad (5.51)$$

We also produced the impulse response for a first order vector autoregression model, with consumption growth (defined as ΔC_t) and expected returns as the endogenous variables.

$$Y_t = A + BY_{t-1} + v_t$$

$$\text{where } Y_t = [\ln C_t \quad \mu_t]', \alpha = [\alpha_1 \quad \alpha_2]', B = \begin{bmatrix} \beta_1 & \beta_2 \\ \beta_3 & \beta_4 \end{bmatrix},$$

$$v_t = \begin{bmatrix} v_{1t} \\ v_{2t} \end{bmatrix}.$$

The above system is estimated and impulse response functions are plotted as a result to the expected returns process. The estimation of (5.51) resulted in a negative relationship between consumption and expected returns. The parameters range from -3.51 to the extreme case of -8.13. No definite distinction between expected returns under the ARFIMA(1,d,0) and AR(1) was found.

The impulse response functions (plots 5.1 and 5.2) show that there is a higher persistence in consumption growth after a shock to the AR(1) expected returns. The expected return series from the ARFIMA model has already accounted for the long memory components and as such, shocks are damped after each lag. This modest finding may be reconciled with business cycle theories, where the frequency of a cycle is shown to be four years.

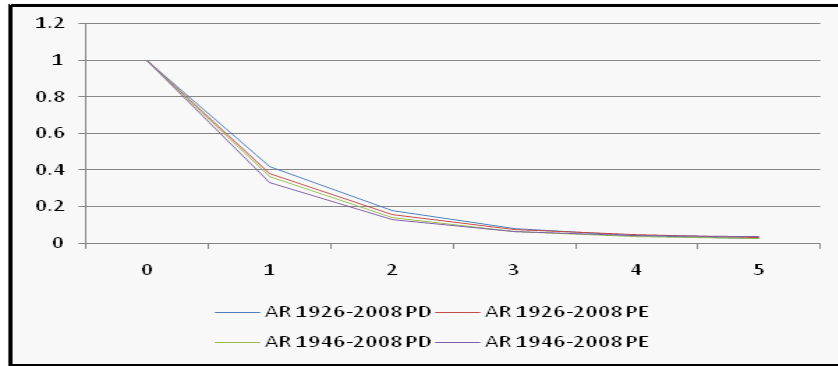


Figure 5.1: **Impulse Response for Consumption Growth - AR(1)**. The table shows the impulse response plots from a shock in the discount rates according to the autoregressive process.

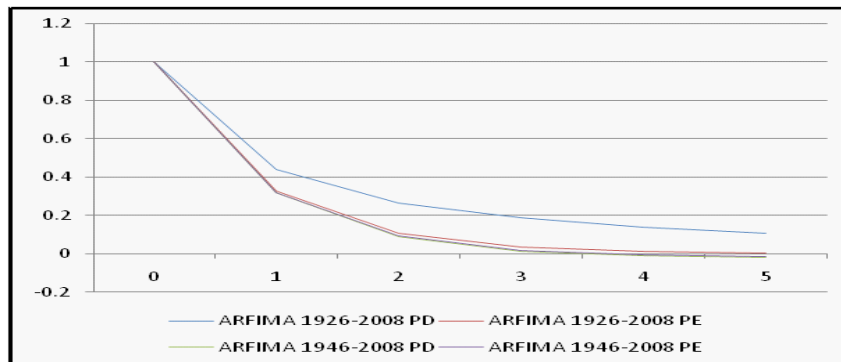


Figure 5.2: **Impulse Response for Consumption Growth - ARFIMA(1,d,0)**. The table shows the impulse response plots from a shock in the discount rates according to the autoregressive process.

Trading Strategy

In this section, an application of the expected returns and expected dividend growth rate series is provided. I test whether the trading strategy may be implemented by identifying whether the stock market is underpriced or overpriced. The trading strategy is an all long strategy with the choice of either going long in bonds or the equity index. If the market is overpriced, reversion towards fundamental value will imply that price will fall in the following period(s), leading to a potential capital loss if equity index is held. In this case, the trading rule is to go long on treasury bills. Likewise, if the equity index is underpriced, reversion to the fundamental value implies that there would be an increase in the price, therefore implying a positive return.

The present value is computed in real-time using the values of the derived series. As a measure of comparison, the present value is computed using the previous period realized values. The trading strategy is compared against the Buy and Hold. Two versions of the present value are assumed. The first model of the present value is the Gordon Dividend Growth model, which assumes that the dividend growth rate is constant. The second present value is the discounted next period dividend and price. The expected future price is proxied by the present price, given the random walk nature of the price.

The present value formulae are computed as follows:

$$PV1^{AR} = \frac{D_t(1 + g_t^{AR})}{\mu_t^{AR} - g_t^{AR}}, \quad (\text{Model 1})$$

$$PV1^{ARFIMA} = \frac{D_t(1 + g_t^{ARFIMA})}{\mu_t^{ARFIMA} - g_t^{ARFIMA}}, \quad (\text{Model 2})$$

$$PV2^{AR} = \frac{D_t(1 + g_t^{AR}) + P_t}{\mu_t^{AR} + 1}, \quad (\text{Model 3})$$

$$PV2^{ARFIMA} = \frac{D_t(1 + g_t^{ARFIMA}) + P_t}{\mu_t^{ARFIMA} + 1}, \quad (\text{Model 4})$$

$$PV^R = \frac{D_t(1 + \Delta d_t) + P_t}{r_{t+1} + 1}. \quad (\text{Model 5})$$

An important point worth mentioning is that in models (Model 3), (Model 3) and (Model 3), the present value contains the lagged price P_t which is used as a proxy for P_{t+1} . However, it should be stressed that a more appropriate model would be P_t with a drift term. Prices are known to be random walk with drift term.⁸

The cumulated returns from the trading strategy (using the present value) are plotted for the different samples and measurement variables in figures A.25, A.26, A.27 and A.28 in the appendix. Based on the four samples, the Buy and Hold strategy tends to beat the trading strategy. The only exception comes from earnings growth for the period 1946-2008 where the present value model 2 tends to beat the Buy and Hold over the whole period. However, it is worth mentioning that the graphical plots do not do justice to the proper performance of the trading strategy since a high market return in one period may bias the Buy and Hold strategy towards having a higher accumulated return than the trading strategy. We report a measure based on the binary outcome of whether the rule advised going towards the highest return (maximum between market returns and bond returns) each year. The percentage of times each model was successful is reported in table 5.6.

⁸The random walk model with drift is written as follows: $P_{t+1} = \mu + P_t + \varepsilon_t$. Therefore the best forecast of $P_{t+1} = \mu + P_t$.

Table 5.6:

Success Rate of Trading Strategy.

The table illustrates the percentage number of times, the strategies posit the correct holding position. The correct holding position is defined as the maximum between the risk free and market return in the corresponding year. The construction of the series is for the different trading strategies are conditional on the different expected returns.

	1926-2008 PD	1926-2008 PE	1946-2008 PD	1946-2008 PE
Model 1	37	53	41	55
Model 2	35	46	40	66
Model 3	36	30	38	46
Model 4	47	57	24	60
Model 5	42	47	46	51
Buy and Hold	69	69	71	71

Interestingly in the binary measure case, none of the trading strategies manage to beat the Buy and Hold. However, compared to the graphical plots the strategies do not perform as badly. There is no definite winner in terms of the present value formulation adopted. Both versions of the present value tend to perform well for the different time periods involved. This finding of non-robustness may be due to the presence of breaks or regime switches, which invalidate the constant dividend growth theory. The ARFIMA models appears to work better than the autoregressive models.

The ARFIMA model, by accounting for hyperbolic decay, is a smoother series, implying that unless there are huge changes in the dividend growth and realized dividends, the present value will smooth over time. In other words, the time series of the present value is itself not volatile. In such a case, it may be likely, that the ARFIMA performs better than the AR models because of its ability to capture smoother business cycle transition over time.

5.6 2 Step Model Estimation Results

In this section, the 2 step procedure is applied to dividend and earnings growth for the sample 1926-2008. The expected returns from the present value is modeled by ARFIMA(0,d,0). First we look at the stationarity of the log linearization parameters. Secondly, we look at the descriptive statistics of the predictive regressions and the properties of the new expected returns.

κ for dividend growth and earnings growth is plotted in figures A.15 and A.16 respectively. κ_t is mean reverting but exhibits higher volatility for earnings growth. With regards to the stationarity of this equilibrium, two observations are made: Firstly, κ_t from the price-dividend ratio is not stationary (p value from ADF is 0.9). However that of earnings growth is stationary (p value of 0.025). The same feature is found present with ρ_t . However, the non-stationarity of dividend growth κ_t does not matter, since it is combined with ρ_t . The resulting linear combination is stationary and ensures that the present value variables are stable over time.

I also model explicitly the expected returns using the pure autoregressive and fractionally integrated model. The results are compared using the standard measures of fit. The summary statistics are illustrated in tables 5.7 and 5.8 for the growth and discount rate models respectively.

Table 5.7:

Summary Statistics of Dividend and Earnings Growth.

The table shows the distributional properties of realized next period dividend and earnings growth and the fitted dividend and earnings growth values. The Mean Squared Error (MSE) is the squared difference between the predictive regression and the realized values.

	Δd_{t+1}	$\bar{g}_{d,t}^{M1}$	$\bar{g}_{d,t}^{M2}$	$\bar{g}_{d,t}^{M3}$	Δe_{t+1}	$\bar{g}_{e,t}^{M1}$	$\bar{g}_{e,t}^{M2}$	$\bar{g}_{e,t}^{M3}$
Mean	0.015	0.012	0.011	0.014	0.007	0.013	0.014	-0.009
Standarddev	0.106	0.002	0.002	0.011	0.228	0.003	0.004	0.143
Jarque-Bera	191	2.884	3.626	4.618	62.98	51.64	54.26	2.714
MeanSquaredError	—	0.011	0.011	0.011	—	0.052	0.052	0.048
A.Dickey-Fuller(p-val)	0.01	0.025	0.01	0.01	0.01	0.1	0.05	0.01

Table 5.8:

Summary Statistics of Dividend and Earnings Growth.

Given the first step once the variables have been fitted in the regression, they are netted off from the present value. The reported statistics are the distributional properties of the expected returns series.

	r_{t+1}	$\bar{\mu}_{d,t}^{M1}$	$\bar{\mu}_{d,t}^{M2}$	$\bar{\mu}_{d,t}^{M3}$	$\bar{\mu}_{e,t}^{M1}$	$\bar{\mu}_{e,t}^{M2}$	$\bar{\mu}_{e,t}^{M3}$
Mean	0.080	0.059	0.057	0.061	0.092	0.093	0.069
Standard dev	0.196	0.003	0.003	0.011	0.021	0.021	0.150
Jarque-Bera	0.140	2.934	2.027	4.053	3.926	3.924	5.972
Mean Squared Error	—	0.039	0.039	0.036	0.039	0.039	0.024
A.Dickey-Fuller (p-val)	0.01	0.9	0.9	0.01	0.9	0.9	0.01

The reported results show that there are considerable differences in the distributions of earnings and dividend growth rate. The predictive regressions have a lower standard deviation than the realized values. All three models for dividend growth tend to marginally under forecast the true dividend growth. On the other hand, earnings growth results are mixed. The present value model tends to forecast a negative earnings growth on average. Models 1 and 2 tend to over forecast. It is found that better forecasts of dividend and earnings growth also leads to a lower mean squared error between expected returns and realized returns.

The tables also illustrate very fascinating results when it comes to the stationarity of the corresponding series. Earnings growth is deemed to be nonstationary at the 5 % level if forecast with the simple mean model. Expected returns is found to be non-stationary when the present value variables are included. This finding are similar to the results reported in tables A.22, A.23, A.24 and A.25 reported in the appendix.

Specification of Expected Returns Process

The same processes adopted in the state space section shall be considered. The expected returns are modeled as an AR(1) and an ARFIMA(0,d,0)⁹. The objective is to compare the fit between the short run process and the long range dependent process. The results are presented in tables 5.9 and 5.10¹⁰.

⁹Variants with an ARFIMA(p,d,q) may also be considered. In this case, the simple AR(1) and ARFIMA(0,d,0) are interesting cases.

¹⁰The specification for the AR(1) and ARFIMA(0,d,) are given by (5.52) and (5.53).

$$\bar{\mu}_t = \alpha \bar{\mu}_{t-1} + \varepsilon_t^{AR} \quad (5.52)$$

$$\bar{\mu}_t = (1 - L)^d \varepsilon_t^{ARFIMA} \quad (5.53)$$

Table 5.9:

**Regression Results for AR(1) and ARFIMA (0,d,0) Specification
using Dividend Growth.**

The table shows the estimates, goodness of fit and misspecification tests if each series was plotted using a univariate framework. The two univariate specifications adopted in this case are the AR(1) and the pure ARFIMA(0,d,0). The state space models are also reported for comparison. Sieve p stands for the bootstrapped p-value from the Sieve-Autoregression.

	2 stage Model						Kalman Filter			
	1926-2008			1946-2008			1926-2008		1946-2008	
	$\bar{\mu}_{d,t}^{M1}$	$\bar{\mu}_{d,t}^{M2}$	$\bar{\mu}_{d,t}^{M3}$	$\bar{\mu}_{d,t}^{M1}$	$\bar{\mu}_{d,t}^{M2}$	$\bar{\mu}_{d,t}^{M3}$	$\bar{\mu}_{d,t}^{AR(1)}$	$\bar{\mu}_{d,t}^{FI(d)}$	$\bar{\mu}_{d,t}^{AR(1)}$	$\bar{\mu}_{d,t}^{FI(d)}$
Intercept	0.06	0.06	0.06	0.06	0.06	0.06	0.04	0.11	0.06	0.06
p-value	0	0	0	0	0	0	0.31	0	0	0
Sieve p	0	0	0	0	0.01	0	0.16	0	0	0.02
α	0.90	0.90	0.35	-	-	-	0.94	-	0.41	-
d	-	-	-	0.95	0.90	0.28	-	0.98	-	0.33
p-value	0	0	0.01	0	0	0.02	0	0	0	0.01
Sieve p	0.10	0.06	0	0.01	0.01	0.01	0	0	0	0
R-Squared	0.79	0.79	0.11	0.79	0.80	0.10	0.87	0.87	0.17	0.17
Log Likelihood	409	412	252	413	416	255	214	215	131	132
Autocorrelation	0.46	0.49	0.80	0.32	0.40	0.90	0.83	0.80	0.65	0.69
Arch Effects	0	0	0.79	0	0	0.80	0.07	0.27	0.02	0

Table 5.10:

**Regression Results for AR(1) and ARFIMA (0,d,0) Specification
using Earnings Growth.**

The table shows univariate estimates from the earnings model. $FI(d)$ refer to the ARFIMA (1,d,0) from the state space model.

	2 stage Model						Kalman Filter			
	1926-2008			1946-2008			1926-2008		1946-2008	
	$\bar{\mu}_{e,t}^{M1}$	$\bar{\mu}_{e,t}^{M2}$	$\bar{\mu}_{e,t}^{M3}$	$\bar{\mu}_{e,t}^{M1}$	$\bar{\mu}_{e,t}^{M2}$	$\bar{\mu}_{e,t}^{M3}$	$\bar{\mu}_{e,t}^{AR(1)}$	$\bar{\mu}_{e,t}^{FI(d)}$	$\bar{\mu}_{e,t}^{AR(1)}$	$\bar{\mu}_{e,t}^{FI(1,d)}$
Intercept	0.09	0.09	0.06	0.09	0.09	0.10	0.07	0.15	0.08	0.08
p-value	0	0	0.10	0	0	0.45	0.14	0	0	0
Sieve p	0	0	0.04	0	0	0.22	0.05	0	0	0
α	0.77	0.77	0.57	—	—	—	0.83	—	0.59	—
d	—	—	—	0.94	0.94	0.52	—	0.96	—	0.53
p-value	0	0	0	0	0	0.02	0	0	0	0
Sieve p	0.07	0.01	0	0.06	0.06	0.02	0.01	0.02	0	0
R-Squared	0.56	0.56	0.31	0.55	0.55	0.27	0.61	0.61	0.36	0.33
Log Likelihood	228	227	54	227	226.	52	131	130	174	174
Autocorrelation	0.03	0.03	0.36	0.01	0.01	0.26	0.06	0.04	0.96	0.90
Arch Effects	0.16	0.16	0.05	0.11	0.11	0.01	0.28	0.16	0.60	0.21

The fit of expected returns is relatively low when dividend and earnings growth are forecast from the present value variables. In these cases, dependence is relatively low as illustrated by the weak autoregressive and fractional parameters. Interestingly, the unconditional expected returns is higher for that case and is deemed not statistically significant using both the conventional p-values and the Sieve-AR wild bootstrap¹¹. Inference on the latter standard errors accounts for the dependence and serial correlation in the variance, as shown by the tests of serial correlation and ARCH effects¹². Compared to the earnings growth model, the expected returns has considerable conditional heteroscedasticity.

The expected returns from the two step approach are compared with those of the state space. State space models which accommodate both short and long range dependence (ARFIMA (1,d,0)) perform very badly with regards to the goodness of fit and the log likelihood. The reason for this weak fit is that the expected returns depends on the truncation level in the state space. Accommodating short run and long run dependence for a small sample such as the present one may be arithmetically daunting. On the other hand, the fit of the AR(1) is very good. It is marginally better than the 2 step regression model according to the R-squared, but fares worse according to the likelihood value. Intuitively the state space model should provide a better fit since both the dynamics and of expected returns and expected dividend (earnings) growth are modeled jointly. In the two step model, no initial assumption is made on the expected returns.

¹¹The Sieve-AR wild bootstrap is used when the residuals are non iid. An interesting introductory explanation is available in Godfrey (2009). For the selection of the skewness parameter in wild bootstrap see Davidson et al (2007). The number of lags for the autoregression parameter is selected using the AIC criterion.

¹²The bootstrapped p-values are consistent even in the presence of serial correlation as reported in numerous instances of our previous results.

5.7 Conclusion

In this chapter, a new parametrization of expected returns is considered. The previous chapters showed that the smoothed expected returns from the state space model may have a unit root. Theoretically, this finding is not possible. One potential reason may be the fact that expected returns was improperly specified as an AR(1) process. If expected returns are assumed to be an ARFIMA process, it means that the latter is decomposed between short and long run dependent components. This ensures that the autoregressive parameter is not contaminated by the long range dependent component. Two ARFIMA models were considered in the case of the present value model. Two sample sizes were considered for both dividend and earnings series. The fitted ARFIMA's tend to show mixed results depending on the methodology (state space and 2 step procedure), variables (earnings and dividend growth) and sample size involved.

In the first case, expected returns was modeled jointly with dividend (earnings) growth in a state space structure. The results from the ARFIMA are not robust across time and it is suspected that regime switches across the two different samples or the small sample size adopted might be the cause. This finding has been backed by a small Monte Carlo experiment with two regimes. The univariate specifications of the expected returns series also show that the AR(1) fares better than the ARFIMA (1,d,0) for the state space. The filtered series was used in three simple applications namely in evaluating return predictability, assessing the relationship between consumption growth and discount rate and lastly in a market timing strategy.

It was found that expected returns had little insample predictability with R-squared ranging from 0.002 to 0.08. Both the AR(1) and

ARFIMA(1,d,0) have equal power in predicting returns, although it is marginally lower than valuation ratios. However the filtered series have stronger forecasting power for dividend and earnings growth rate. Consumption and expected returns was found to be negatively related and a simple impulse response function showed that the effect of a shock in the discount rate may last until four years on consumption growth. The results on the trading rule is that it is impossible to jointly build a mean reverting strategy by identify the over or under pricing of the equity market exante. Such a strategy performs poorly against the Buy and Hold.

Tests of persistence and time variation were performed on the different series. Expected returns was found to be both persistent and time varying. Expected dividend growth was found to be only time-varying. When the ARFIMA model was adopted in an univariate framework, tests of nonstationarity showed that the series was stationary. The two stage expected returns model showed better fit of the ARFIMA model. Formal tests of long memory will be performed in the next chapter. This chapter contributes to the existing literature in that it reconciles the empirical findings of Kojien and Van Binsbergen (2011) with the theory that expected returns must be stationary. A distinction between short and long range dependent components was made. The filtered expected returns series ensures that the series is stationary.

Chapter 6

Test of Long Memory based on Self-Similarity

6.1 Introduction

Expected Returns may exhibit long memory, as shown in the previous chapter. However, the previous chapter compared the ARFIMA structures with AR structures. The models were compared based on goodness of fit and univariate specifications. However, it is possible that expected returns may follow a different kind of process including nonlinear alternatives. In order to whether the process is indeed a long memory process, a skip-sampled test is developed where the null hypothesis is that the series is indeed a long memory. This chapter introduces the notion of long memory and explains its asymptotic properties. One of the problems encountered when observations are sampled from the observed series is the problem of aliasing. The paper explains how the aliasing induced bias is corrected. The size and power of the test are investigated by Monte Carlo. The test is applied to individual stocks and to the series of expected returns.

Long memory is a feature exhibited by many financial series.

Compared to other processes, it consists of hyperbolically decaying autocorrelation and impulse weights. In this case, the process is no longer $I(0)$ or $I(1)$ but is said to be fractionally integrated $I(d)$. Long memory may be defined in the time or frequency domain. In the time domain, it simply means that the sum of the correlations over time is nonfinite. In the frequency domain, a process is said to exhibit long memory if the spectral density function is unbounded at low frequencies. The test which is developed is semi parametric in nature and is developed within the frequency domain.

Long memory may turn out to be spurious in the presence of structural breaks and nonlinearities (Diebold and Inoue 2001). It is important to test for true memory. Various tests have been used in the literature for the estimation and testing of ‘ d ’. Tests of $I(0)$ versus an $I(d)$ include the Harris, Leybourne and McCabe test (2008), Lo’s Modified Range Scale test (1991), Rescaled variance test Giraitis et. al. (2003) Other parametric test include the test specification from Robinson (1995) and Velasco (1999), which involves doing some pretesting to check if ‘ d ’ is long memory with finite variance ($0 \leq d \leq 0.5$). In the case of tests of $I(1)$ against fractional alternatives, Delgado and Velasco (2005) present a test of long memory based on the sign of the residual terms. It allows for various nulls namely, one when the null is nonstationary and the other when the series is weakly dependent. Dolado, Gonzalo and Mayoral (2005) develop a test of $I(1)$ against fractional alternatives which is in the same spirit as the Dickey Fuller test. It was later refined by Lobato and Velasco (2007) who adapted the test to the case of an unidentified d .

The application of bootstrap models to test for long memory is recent. Grau-Carles (2006) tests four estimators of fractional integration namely the range scale statistic, the modified range scale

and the GPH estimator and the detrended fluctuation statistic. the modified range scale and GPH estimator were the best models. Bootstrapping has also been used to test for fractional cointegration (Davidson 2002a, 2002b, 2004). The application of the bootstrap requires that the autocorrelation structure of the stochastic realizations are maintained. in this spirit, moving block bootstraps may be used. In this study, the Sieve Autoregressive bootstrap is used.

6.2 Long Memory

Semiparametric estimation of long memory is a popular methodology in time series analysis. When the autocovariances of a process are nonsummable, the spectral density f diverges at the origin with

$$f(\lambda) = O(|\lambda|^{-2d}) \quad (6.1)$$

as $\lambda \rightarrow 0$ and this characteristic provides the basis for semiparametric estimation of the long memory parameter d , which is zero in the summable case. The log-periodogram regression originally proposed by Geweke and Porter Hudak (1983, henceforth GPH) is a popular implementation of this idea. The convenient linear fractional differencing representation of long memory (Granger and Joyeux 1980, Hosking 1981) in which

$$f(\lambda) = |1 - e^{-i\lambda}|^{-2d} g(\lambda) \quad (6.2)$$

where g is bounded at the origin, is often adopted, and since $|1 - e^{-i\lambda}|^2 = \lambda^2 + O(\lambda^4)$ there is little loss of generality in assuming (6.2) provided g is otherwise unrestricted.

An inherent problem with this approach to investigating memory characteristics is that there are time series models for which the

GPH estimator will return large and significant values of d in finite samples, in spite of the fact that the autocovariances are summable. This is due to the well-known finite-sample bias of the estimator due to the neglect of short-run dependence components. The simplest illustration of the difficulty is provided by the observational equivalence between the fractionally integrated process $(1 - L)^d x_t = u_t$ with $d = 1$ and the autoregressive process $(1 - \phi L)x_t = u_t$ with $\phi = 1$. For every finite sample size, there exists a ϕ in the latter model large enough to bias the GPH estimator of d significantly, when its true value is zero. It is therefore desirable to have a means of distinguishing the cases of true d and spurious d .

Recent research has pointed to the well-known property of self-similarity of hyperbolic decay processes under rescaling transformations, such as periodic aggregation and periodic sub-sampling otherwise known as skip-sampling. Chambers (1998) was the first to point out that if a long memory process is recorded at different rates, the rate of decay of the autocovariances is invariant to the rate of observation. There are two ways to conceive of lowering the observation rate. Temporal aggregation means taking the sums of n successive observations to create the new sequence. This is the natural transformation in the context of flow data, such that (for example) quarterly flows are each the sum of three successive monthly flows. Ohanissian et. al. (2008) implement a test of long memory based on comparing different rates of temporal aggregation.

Skip-sampling, by contrast, means taking every n th observation and discarding the remainder. This is the natural way of lowering the observation rate for stock or price data, although for the present purpose the nature of the observations is irrelevant, since the required properties of the skip-sampled series hold in all cases. Consider these in the context of hyperbolic memory decay. Let the

parameter δ index the rate of decay such that the autocovariance sequence satisfies

$$\gamma_j = O(j^{-\delta}) \quad (6.3)$$

for some $\delta > 0$. The hyperbolic memory class includes short memory processes having summable autocovariances, such that $\delta > 1$, and the long memory class where $\delta = 1 - 2d$. It is immediately evident that, for any fixed, finite n ,

$$\gamma_{nj} = O(j^{-\delta}).$$

It follows that for the long memory class, the property of the spectral density at the origin should likewise be invariant to the sampling frequency. This is in contrast to the case of exponential memory decay, where $\gamma_j = o(j^{-\delta})$ for every finite δ , but there exists $\rho > 0$ such that

$$\gamma_j = O(e^{-\rho j}). \quad (6.4)$$

In this case, note that

$$\gamma_{nj} = O(e^{-\rho nj})$$

so that the memory decay parameter rises from ρ to ρn . Since the estimator of (spurious) d in the exponential decay case is likely to be sensitive to the value of ρ , this suggests that comparing estimates with different rates of sampling might yield a useful test of the null hypothesis of long memory.

A range of nonlinear models, such as ESTAR, SETAR and Markov-switching autoregressions are often thought of as likely to be mistaken for long memory, since they can exhibit local patterns of apparent persistence, switches of local mean, for example, or unit root-like behaviour in the neighbourhood of the origin. As in the case of the linear autoregressive model, the essential difference be-

tween these latter models and the long-memory case is that the serial dependence decays exponentially as the lag increases beyond a certain point, whereas long memory implies hyperbolic decay. Whether linear or nonlinear, stable difference equations of finite order necessarily exhibit exponential decay (see Gallant and White 1988, Davidson 1994), whereas unstable difference equations feature. $\delta = 0$.

Note that the class of cases of (6.3) with $1 \leq \delta < \infty$ count as instances of the alternative for present purposes. In any case, models of this sort do not seem to have been significantly exploited to date, except in the rather special contexts of over-differenced fractional models (where $d < 0$) and stochastic volatility modelling, the FIGARCH (Baillie et. al. 1996) and HYGARCH (Davidson 2005) models being cases of the ARCH(∞) model where the lag weights in the conditional variance equation decline hyperbolically but are nonetheless summable.

This paper considers tests of the long memory hypothesis based on a comparison of the log-periodogram estimator of the d parameter in skip-sampled data with that from the original data. The test statistic we have in mind, the simple difference of the two estimators, is asymptotically Gaussian under the usual assumptions of this literature (notably, Gaussianity of the observations) and is also asymptotically pivotal. However, we have not attempted to derive the limiting variance, since there are various reasons to suppose that the convergence to the limit is slow, and this is in any case a natural application for a bootstrap test. We develop bootstrap procedure that should be asymptotically correctly sized for processes that are linear under the null hypothesis, at least. The rest of the paper is organized as follows. Section 2 reviews the important issue of aliasing and its consequences for the form of the periodogram. Section 3 describes the bootstrap test, Section 4 provides additional details on

the aliasing correction that has been adopted for the skip-sampled estimator, and Section 4 discusses the null asymptotic distribution of the statistic. Section 5 discusses the appropriate usage of the test and its properties under the alternative. Our Monte Carlo findings are reported in Section 6.

6.3 Aliasing

The distribution of the GPH estimator in skip-sampled data has been studied *inter alia* by Chambers (1998), Smith and Souza (2002) and Souza (2003, 2005). Skip-sampling induces a bias in the estimator due to the aliasing effect. For a comprehensive analysis of the aliasing phenomenon, see Hassler (2011). The basic result is that the spectral density of the skip-sampled data can be represented as an average of the spectral densities over the range of aliased frequencies.

Proposition 1 *If $\{x_t, t = 1, 2, \dots\}$ is a discrete stochastic process with spectral density f and $y_t = x_{nt}$ for $t = 1, 2, \dots$, the spectral density of the process y_t is*

$$f_n(\lambda) = \frac{1}{n} \sum_{j=0}^{n-1} f\left(\frac{\lambda + 2\pi j}{n}\right).$$

The straightforward proof is given in the appendix. Note that cycles of frequency λ/n in the original data become cycles of frequency λ in the skip-sampled data, and frequencies higher than π/n are no longer identifiable. Hence, these contributions to the variance of the series are effectively aggregated with the identifiable frequencies.

Apply this formula to the case

$$f(\lambda) = |1 - e^{i\lambda}|^{-2d} g(\lambda) = [2 \sin(\lambda/2)]^{-2d} g(\lambda), \quad (6.5)$$

where $0 < g(0) < \infty$, $g'(0) = 0$, $g''(0) < \infty$. We find that $f_n(\lambda)$ unfortunately does not admit to direct log-linearization in the GPH manner. What can be done, however, following the suggestion of Smith and Souza (2002), is to write

$$\begin{aligned} f_n(\lambda) &= \frac{1}{n} \sum_{k=0}^{n-1} \left(2 \sin \left(\frac{\lambda + 2\pi k}{2n} \right) \right)^{-2d} g \left(\frac{\lambda + 2\pi k}{n} \right) \quad (6.6) \\ &= \left(2 \sin \frac{\lambda}{2n} \right)^{-2d} g \left(\frac{\lambda}{n} \right) H_n(\lambda), \end{aligned}$$

where

$$H_n(\lambda) = \frac{1}{n} \sum_{k=0}^{n-1} \left(\frac{\cos(\lambda/2n)}{\sin(\lambda/2n)} \sin(\pi k/n) + \cos(\pi k/n) \right)^{-2d} \frac{g((\lambda + 2\pi k)/n)}{g(\lambda/n)}. \quad (6.7)$$

There is evidently an additional omitted term $\log H_n$ in the log-periodogram regression, depending on d as well as λ . Its coefficient is known to be unity, but its omission is not rendered negligible by taking frequencies close to the origin. The omission of this term will be liable to produce a bias in the GPH regression. What is commonly observed is that estimates of $d > 0$ obtained from skip-sampled data are substantially closer to zero than those from the original data.

Remark 2 *Note the implication for the standard analysis of a model such as 6.5, which is revealed to be specifically linked to the frequency of the observation. Without this assumption, there is no reason to suppose that the function g does not depend on d , nor that it is constant near the origin. In this light, the standard long memory analysis appears a little more fragile than is commonly taken for granted. Nonetheless, in this paper, these assumptions shall be used for the purposes of developing a test.*

6.4 The test

The test we propose is based on the comparison of two estimators of the memory parameter d , one based on the full sample, the other based on periodic sampling, otherwise called skip-sampling, of the test series. We let n denote the periodicity of the sample. In principle, any one of a number of different estimators might be adopted to implement the test, but in this paper we adopt the Geweke Porter-Hudak (1983) (GPH) log-periodogram regression estimator, which has the obvious benefit of ease of calculation .

Skip-sampling is done taking every n th observation, so yielding a sample size $[T/n]$, where $[z]$ denotes the largest integer less than z . this can be done n times by off-setting the initial observation, so that the n skip-samples are $\{x_{0t}\}, \{x_{1t}\}, \dots, \{x_{n-1,t}\}$ where $k = 0, \dots, n - 1$

The test has been implemented as a bootstrap test, where the distribution under the null hypothesis is simulated as a fractionally integrated process, although allowing for the possibility of short-run dependence of the fractional differences. The steps leading to the computation of a p -value for comparison with the chosen significance level are as follows.

1. Let the conventional GPH estimator, based on the complete sample be denoted as \hat{d} and let modified log-periodogram estimators, with a bias correction to be explained in greater detail in the next section , be denoted \hat{d}_{nk} for $k = 0, \dots, n - 1$. The test statistic is defined as

$$\tau = \hat{d} - \hat{d}_n, \quad (6.8)$$

where $\hat{d}_n = n^{-1} \sum_{k=0}^{n-1} \hat{d}_{nk}$.

The test is implemented as a bootstrap test where the distribution under the null hypothesis is simulated as a fractionally

integrated process, allowing for the possibility of linear short-run dependence of the fractional differences. Thus linearity is the chief restriction on the class of models included in the null hypothesis. The sieve-autoregression procedure (Bulmann 1997) is used to model this dependence in the bootstrap draws. Given \hat{d} and τ as in 6.8, computed from the observed sample, the steps leading to the computation of a p-value for comparison with the chosen significance level are as follows

2. Compute the fractional differences $\hat{u}_t = (1 - L)^{\hat{d}}y_t$.
3. Fit an autoregression of order p_T for \hat{u}_t , using the Durbin-Levinson algorithm, where p_T is chosen to optimize the Akaike criterion subject to $p \leq 0.6T^{1/3}$. Let $\hat{\varepsilon}_t = \hat{\phi}(L)\hat{u}_t$ denote the residuals from this model.
4. Repeat the following steps for $j = 1, \dots, B$
 - (a) Draw a sample $\hat{\varepsilon}_{1j}^*, \dots, \hat{\varepsilon}_{Tj}^*$ with replacement from the distribution $P(\hat{\varepsilon}_t^* = \hat{\varepsilon}_t) = 1/T$, and
 - (b) Generate the bootstrap data sample as

$$\hat{y}_{tj}^* = (1 - L)^{-\hat{d}}\hat{\phi}(L)^{-1}\hat{\varepsilon}_{tj}^*1_{\{t \geq 1\}} + \widehat{z}_{tj}, \quad t = 1, \dots, T$$

where $\hat{\varepsilon}_{tj}^*1_{\{t \geq 1\}} = 0$ for $t < 1$ but the sequence $\widehat{z}_{1j}, \dots, \widehat{z}_{1j}$

- (c) Compute the bootstrap statistic τ_j^* for the sample $\hat{y}_{1j}^*, \dots, \hat{y}_{Tj}^*$.

5. Compute the estimated p-value for the test as

$$1 - \frac{\min\{j : \tau \leq \tau_{(j)}^*\}}{B},$$

where $\tau_{(j)}^*$ is the j th order statistic for the bootstrap statistics $\tau_1^*, \dots, \tau_B^*$.

Remarks

1. We use the signed test statistics and hence do a one-tailed test, on the assumption that the leading cases of the alternative will give rise to a smaller value of d in the skip-sampled data.
2. The use of the average d estimates from the n skip samples available makes the most efficient use of the available data. In view of the form of the estimator, this is equivalent to adopting the average of the log-periodogram points as regressand.
3. The correction terms \widehat{z}_{jt} are constructed using Gaussian drawings and weights constructed from the estimated parameters to have a covariance structure matching the components omitted through the truncating of the sequence at 0. The resulting sequence is approximately stationary for $|d| < 0.5$. If $\hat{d} \geq 0.5$ the data are modelled in differences, replacing \hat{d} by $\hat{d} - 1$, and the simulation is then integrated using the first observation for the initial condition. Note that nonstationary processes generated by this procedure converge after normalization to Type 1 Brownian motion. For details of the simulation procedure, see Davidson and Hashimzade (2009).

6.5 Bias Correction

The construction of the estimators \hat{d}_{nk} used in the test is a key issue. As shown in section 2 the conventional GPH estimator applied to skip-sampled data is biased. The magnitude of the bias depends on d as well as the form of the short-run dependence, and this bias is not attenuated by a narrow bandwidth. Since the test is to be implemented with the bootstrap, bias might not be regarded as a critical issue here. An asymptotically correctly sized test is assured provided the bootstrap replications can reproduce the distribution of τ under the null hypothesis, and having the mean of this statistic different from zero under the null hypothesis would apparently not

be a critical matter. However, there are two main reasons why this aspect of the procedure is in fact crucial.

The first reason is that the dependence of the mean of the statistic on nuisance parameters implies that the null distribution is not asymptotically pivotal. As is well known (see e.g. Horowitz 2000) this will have the effect of increasing the order of magnitude in T of the error in rejection probability (ERP). This is especially important because of the relatively slow rates of convergence of the narrow band estimators employed here. The bias-corrected test has mean zero asymptotically, still exhibiting some dependence on the unknown d , but of $O_p(1)$.

Second, the test is designed to determine whether "large" (nominally significant) estimates of the d parameter are spurious or consistent for the rate of hyperbolic memory. The power of the test depends on the skip-sampling estimates lying "significantly" closer to zero under alternatives than the full-sample estimates. When the null is false, however, the long-memory component of the bias term that would be present under the null hypothesis is absent; hence the test comparison becomes a matter of comparing one bias with another. We conjecture that test power could be correspondingly poor.

Bias correction involves finding a computable surrogate for H_n in (6.7), and there are two issues to be considered. The expression depends in the first place on the unknown parameter d , and in the second place on the spectral density component $g(\lambda)$. For the purposes of the test, estimates from the skip samples are to be compared with those from the full sample. Therefore, the natural approximation is to replace d with the asymptotically unbiased estimator \hat{d} .

Except in the case of the pure fractional difference, the term $g(\lambda + 2\pi j/n) / g(\lambda/2)$ will in general vary with λ over the whole of the interval $[0, 2\pi]$, including points close to the origin. Approximating it by a constant is therefore not an attractive option, notwithstanding that this is the approach for dealing with $g(\lambda/2)$ in the narrow band estimator. A more ambitious option is therefore to model g , parametrically or semiparametrically. In view of the fact that the sieve-autoregression is to be used to simulate the data for the purposes of the bootstrap test, a natural approach is to make use of these fitted parameters, and so approximate g by

$$\hat{g}(\lambda) = |\hat{\phi}(e^{-i\lambda})|^{-2}, \quad (6.9)$$

where, as above, the autoregressive parameters $\hat{\phi}(L)$ are estimated using the Durbin-Levinson algorithm from $\hat{u}_t = (1-L)^{\hat{d}}x_t$. Essentially this option trades the requirement to assume a linear process under the null hypothesis for the greater efficiency afforded by the autoregressive parameterization.

Letting $\lambda_j = 2\pi j/T$ as usual, the skip-sampled series consists of $[T/n]$ observations, and the frequencies at which the periodogram is evaluated are $\lambda_{nj} = 2\pi nj/T$ for $j = 1, \dots, M_n$ where $M_n = [(T/n)^q]$, for $0 < q < 1$, represents the usual GPH bandwidth function of sample size. Let I_{nk} denote a periodogram computed from k th skip-sampled data and let $\hat{H}_n(\lambda)$ denote formula in (6.7) approximated as described, using the estimated parameters and the representation of the short-run spectral density in 6.7. The bias-corrected skip-sample estimator then take the form:

$$\hat{d}_{nk} = \frac{\sum_{j=1}^{M_n} (X_{nj} - \bar{X}_n) [\log I_{nk}(\lambda_{nj}) - \log \hat{H}_n(\lambda_{nj})]}{\sum_{j=1}^{M_n} (X_{nj} - \bar{X}_n)^2}, \quad (6.10)$$

where $X_{nj} = -2 \log(2 \sin \lambda_{nj}/2)$. Provided n is treated as fixed and

not linked to sample size note that $M_n = O(M)$ where $M = [T^q]$ and this is the assumption we maintain henceforth.

6.6 Asymptotic distribution of the statistic

Let the null hypothesis specify that the random sequence is stationary and Gaussian, having a Wold representation of the form

$$x_t = (1 - L)^{-d}\theta(L)\varepsilon_t,$$

where $\theta(L)$ is an invertible lag polynomial of potentially infinite order and $\varepsilon_t \sim NID(0, \sigma^2)$. Note that since $\theta(L)$ is arbitrary apart from having summable coefficients, this representation does not actually impose a fractionally differenced structure on the data. Every linear Gaussian process having non-summable autocovariances satisfying $\gamma_j = O(j^{2d-1})$ can be given this representation.

When the sample is large enough, both the conventional GPH estimator \hat{d} and the skip-sampled estimator \hat{d}_n defined in (6.10) can be analysed using the techniques developed in Hurvich Deo and Brodsky (1998) (henceforth HDB). In other words, letting $\varepsilon_{nkj} = \log(I_{nk}(\lambda_{nj})/f(\lambda_{nj}))$ there exists a function f^* such that (reproducing the expression in HDB page 42)

$$M_n^{1/2}(\hat{d}_{nk} - d) = -\frac{M_n^{1/2}}{2S_n} \sum_{j=1}^{M_n} (X_{nj} - \bar{X}_n) \log f_{nj}^* - \frac{M_n}{2S_n} \frac{1}{M_n^{1/2}} \sum_{j=1}^{M_n} (X_{nj} - \bar{X}_n) \varepsilon_{nkj},$$

where $S_n = \sum_{j=1}^{M_n} (X_{nj} - \bar{X}_n)^2 = O(M)$, and the first right-hand side term is $o(1)$ on the conditions $M \rightarrow \infty$ and $(M \log M)/T \rightarrow 0$. The case $n = 1$, $k = 0$ is the standard case, without skip-sampling, so that $f_{nj}^* = f_j^* = g(\lambda_j)$, and $\varepsilon_{nkj} = \varepsilon_j$, while the case $n > 1$ has

$$\log f_{nj}^* = \log g\left(\frac{\lambda_j}{n}\right) - \log\left(\frac{\hat{H}(\lambda_{nj})}{H(\lambda_{nj})}\right).$$

Recalling that \hat{d} is $M^{1/2}$ -consistent, and noting that H is twice-differentiable with respect to d , expand $\log \hat{H}(\lambda_{nj})$ as

$$\log \hat{H}(\lambda_{nj}) = \log H(\lambda_{nj}) + \frac{H(\lambda_{nj})'}{H(\lambda_{nj})}(\hat{d} - d) + O(M^{-1}).$$

Then, using Lemma 1 of HDB, and letting

$$B_T(n, d) = \frac{1}{2S_n} \sum_{j=1}^{M_n} (X_{nj} - \bar{X}_n) \frac{H(\lambda_{nj})'}{H(\lambda_{nj})},$$

we have

$$M_n^{1/2}(\hat{d}_{nk} - d) = n^{-q/2} B_T(n, d) M^{1/2}(\hat{d} - d) - \frac{M_n}{2S_n} \frac{1}{M_n^{1/2}} \sum_{j=1}^{M_n} (X_{nj} - \bar{X}_n) \varepsilon_{nkj} + o(1).$$

Be careful to note that the relevant properties of the ε_{nkj} extend to the skip-sampled case; specifically, that their distribution has finite second moments that asymptotically do not depend on nuisance parameters – see Lemmas 2 and 6-8 of HDB. These random variables represent continuous transformations of the discrete Fourier transforms of the data, which in the skip-sampled case are simply weighted sums of the original data points where the weights are zeros except for every n th observation and otherwise are the usual trigonometric functions, defined on the interval $[0, \pi/n]$. Since the regressors are the same for each k , we further find

$$M^{1/2}(\hat{d}_n - d) = B_T(n, d) M^{1/2}(\hat{d} - d) - \frac{n^{-q/2-1} M}{2S_n} \frac{1}{M^{1/2}} \sum_{j=1}^{M_n} (X_{nj} - \bar{X}_n) \sum_{k=0}^{n-1} \varepsilon_{nkj} + o(1). \quad (6.11)$$

In the appendix, we show the following

Proposition 3 *$B_T(n, d)$ converges in probability to a finite non-stochastic limit $B(n, d)$.*

It follows that apart from terms of small order, $M^{1/2}(\hat{d}_n - \hat{d})$ is

asymptotically Gaussian with a mean of zero and finite variance. However, the distribution is not free of nuisance parameters since it depends on n as well as on $B(n, d)$. These dependencies warrant the use of the bootstrap, as the most practical implementation of the test.

6.7 Properties under the alternative hypothesis

Testing the degree of persistence of time series is a testing problem that has attracted a degree of controversy, as has been documented by one of the present authors (Davidson 2009). This is one of a class of problems have been characterized by Dufour (1997) as "ill-posed", and has close links with the testing frameworks critically by Potscher (2002) and Faust (1996,1999), *inter alia*. Tests of the null hypothesis that the series has summable autocovariances – the "I(0) hypothesis" – face a common difficulty for valid inference. This difficulty manifests itself in different ways in different contexts, but the essential common problem might be summarized as follows: cases of the null hypothesis constitute an open set in the parameter space, and leading cases of the alternative are contained in the closure of that set. It follows that test power cannot exceed test size, where the latter is defined as the supremum of the rejection probabilities over the null set of model space.

While this problem extends to much more general parameterizations of the null, it is most transparent in the case where the "I(0)" property depends on the modulus of the maximal autoregressive root, and the null hypothesis is represented by the interval $[0,1)$. The present case is clearly similar, except that the null hypothesis, relating to the value of d , is the case of the open interval $(0, \infty)$, with its closure containing the alternative case $d = 0$. This is another situation where, under a literal interpretation, power cannot exceed

size. For this reason, it is important to emphasize the context in which this type of test might be useful.

The test is based on a comparison of two estimators of d , where under the alternative, one (the full sample estimator) is expected exhibit more bias than the other (the skip-sampled estimator), *as an estimator of zero*. Since the estimators being compared are both consistent, albeit biased in finite samples, the test appears inconsistent. In the limit as the sample size tends to infinity, the null distribution simulated in the bootstrap test is converging on a case of the alternative (short memory) as the bias in both estimators of d converges on zero. There is no reason to suppose that the probability of exceeding the rejection criteria is increasing in sample size.

However, suppose that the test is treated as the second stage of a two-stage procedure, being performed only in the case of rejection in a significance test on d . The null hypothesis to be regarded as rejected, in the case of a non-rejection in this first stage test. This combined test is clearly consistent. It is not feasible to compute the exact significance level of this two-step test, of course, but we may bound it as follows. Let β_S denote the probability of type 2 error at the first stage, incorrectly finding insignificance when the null hypothesis of long memory is true. If the long memory test is conducted at α , but only if there is a first stage rejection, then

$$\begin{aligned} P(\text{reject on composite test} \mid \text{true}) &\leq \beta_S + \alpha(1 - \beta_S) \\ &= \alpha + \beta_S(1 - \alpha) \end{aligned}$$

where the right hand bound exceeds α , but approaches α as β_S approaches 0. Since the first stage is consistent, we can conclude that the combined test is correctly sized asymptotically. Some evidence on this procedure is given in the following section.

An issue not so far addressed is the status as cases of the null hypothesis of nonstationary fractional processes having $0.5 \leq d \leq 1$. It is known (form, e.g. Velasco 1999, Kim and Phillips 2006) that log -periodogram regression in this range is consistent, and also asymptotically normal, under regularity conditions, for $d < 0.75$. Our experiments, reported in the following section, report results for both stationary and nonstationary cases of the null, with similar results. It is noteworthy on this context that autoregressive generated series with a root close to unity characteristically yield an estimated d in the non-stationary range. This is, of course, what the observational equivalence issue raised in the Model section would lead us to predict. A unit root in the case of the null hypothesis, noting that this case exhibits the invariance of the memory to skip-sampling characteristic of the fractional integration case. Considering a sequence of models with maximal autoregressive modulus ranging from unity down to zero, we expect (with fixed sample size) to find the nominal rejection probabilities initially increasing over this range. They may not vary monotonically over the range, but by interpreting non-rejection in the significance test on d as the evidence for the alternative, we are able to discount the behaviour of the test in cases of d close to zero.

In a well-known paper, Diebold and Inoue (2001) point out that in certain models exhibiting structural change in which the frequency of change has a particular relation with sample size, there is the "appearance" of the hyperbolic memory decay. In some of their examples, the processes in question are "revealed" as $I(1)$ as T is extended with fixed parameters. As pointed out, the present test is not expected to have power in such cases. In essence a skip-sampled unit root process remains a unit root. However, there are also examples where the processes are "revealed" as $I(0)$, and in particular these authors consider a simple independent process sub-

ject to Markov Switching. This is one of the case we consider in simulation experiments in the next section.

6.8 Monte Carlo Experiments

Each of the tables in this section shows the results of experiments with four sample sizes (T) two choices of GPH bandwidth M , expressed as fractional powers of T , and three alternative skip rates. In each case, 5000 replications have been performed. The tables show the fifth percentiles of the distribution of the bootstrap p-values.

In this framework, the simulation evidence we present is of three types. First, cases of the null hypothesis are generated as ARFIMA (1,d,0) models, with the form

$$(1 - \phi L)(1 - L)^d y_t = \Theta(L)\varepsilon_t,$$

where here, and also in all the subsequent cases reported, $\varepsilon_t \sim N(0, 1)$. Table 6.1 shows two cases stationary and nonstationary of the pure fractional model with $\phi = 0$, while table (?) reports the case of $d=0.4$ and three cases of the autoregressive parameter ϕ .

The size distortion here, apart from experimental error, results from the failure of the bootstrap data generation procedure to reproduce the distribution of the observations. The primary source of such distortion is bias in the estimator of d , which is of course smaller when the bandwidth is chosen smaller. Since it is bias in the estimator that is also the basis for detecting spurious long memory, there is inevitably a trade off of power for size in the choice of the bandwidth. Taking the case of M in the vicinity $T^{0.7}$ will be the most advantageous from the point of view of power, and happily

the size distortion appears moderate unless the autoregressive component is large. It is also to be noted from the right-hand columns of (?) that the size distortion is minimized when the skip -rate is larger, and this holds for all sample sizes. The reason for this finding is not immediately clear.

Table 6.1:

Pure Fractional Process.

The table shows the power of the test under the null hypothesis of a pure fractional for different sample sizes with skip sampling levels $n = 4, 8$ and 12 . The null hypothesis includes both a stationary ($d = 0.4$) and a nonstationary ($d = 0.6$) fractional parameters. The significance level reported is 5%.

d	T	$M = [T^{0.55}]$			$M = [T^{0.7}]$		
		$n = 4$	$n = 8$	$n = 12$	$n = 4$	$n = 8$	$n = 12$
0.4	250	0.047	0.047	0.048	0.057	0.044	–
	500	0.046	0.045	0.047	0.051	0.057	0.045
	1000	0.051	0.051	0.047	0.042	0.045	0.043
	2000	0.051	0.065	0.052	0.053	0.049	0.053
0.6	250	0.052	0.046	0.048	0.048	0.042	–
	500	0.049	0.048	0.051	0.051	0.044	0.047
	1000	0.045	0.048	0.044	0.049	0.049	0.050
	2000	0.049	0.049	0.047	0.048	0.048	0.044

In tables 6.3 and 6.4, we consider cases of the alternative cases of the alternative hypothesis. Note that these are not size-corrected powers, which would be impossible to construct (see the remarks of the last section) but merely the relative frequencies of rejection at the 5 %level. These tables include the average values of the GPH estimates of d from each experiment, to put the results into context in terms of the findings in a conventional analysis of the data when the true d is zero.

Table 6.3 shows the results for the linear AR(1) case, where the model is

Table 6.2:

ARFIMA(1,d,0) Models.

The table shows the rejection probabilities at the 5 % level for the ARFIMA(1,d,0) model. The autoregressive parameter is set to 0.3, 0.45 and 0.7. The fractional parameter (d) is set to 0.4.

ϕ	T	$M = \lceil T^{0.55} \rceil$			$M = \lceil T^{0.7} \rceil$		
		$n = 4$	$n = 8$	$n = 12$	$n = 4$	$n = 8$	$n = 12$
0.3	250	0.048	0.050	0.045	0.083	0.059	–
	500	0.051	0.048	0.047	0.081	0.060	0.062
	1000	0.046	0.051	0.066	0.070	0.067	0.053
	2000	0.047	0.049	0.050	0.063	0.088	0.062
0.45	250	0.049	0.054	0.049	0.113	0.101	–
	500	0.053	0.048	0.055	0.120	0.085	0.081
	1000	0.048	0.049	0.051	0.112	0.073	0.091
	2000	0.049	0.052	0.050	0.094	0.243	0.076
0.7	250	0.079	0.074	0.072	0.293	0.305	–
	500	0.075	0.065	0.059	0.375	0.312	0.224
	1000	0.063	0.061	0.056	0.401	0.305	0.295
	2000	0.046	0.046	0.053	0.385	0.305	0.287

$$y_t = \phi_1 y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(0, 1)$$

for three alternative values of ϕ_1 . Beyond this familiar class, our problem is to deal with the profusion of possible alternatives, and the cases we report are necessarily chosen rather than arbitrarily, although sharing the characteristic that the values of d obtained in the log-periodogram regression are not too close either to zero or to unity.

- "Bilinear" is a model of the form

$$y_t = \phi_1 y_{t-1} + \phi_2 (y_{t-2} \cdot v_{t-1}) + \phi_3 (y_{t-1} \cdot v_{t-2}) + \varepsilon_t,$$

with $\phi_1 = 0.8$ and $\phi_2 = \phi_3 = 0.3$.

- SETAR is the "self-exciting threshold AR" case

$$y_t + (\alpha_{11} + \alpha_{12}y_{t-1}) * (G_t) + (\alpha_{21} + \alpha_{21}y_{t-1})(1 - G_t) + \varepsilon_t,$$

where $\alpha_{11} = 1, \alpha_{12} = 0.45, \alpha_{21} = 1, \alpha_{22} = 0.9$

$$G_t = \frac{1}{1 - e^{-\gamma(y_{t-1} - y^*)}},$$

with $\gamma = 10$ and $y^* = 1$.

- ESTAR is the "exponential threshold AR" case

$$y_t = \alpha_1 y_{t-1} (1 - e^{(-\gamma y_{t-1}^2)}) + \alpha_2 y_{t-1} + \varepsilon_t,$$

where $\alpha_1 = -1.5, \alpha_2 = 1, \gamma = 0.01$

- "Markov" is a model with Markov-switching intercepts. This model takes the form

$$y_t = \alpha(S_t) + \varepsilon_t,$$

where $S_t = 1$ or 2 , and $P(S_t = 1 | S_{t-1} = 2) = P(S_t = 2 | S_{t-1} = 1) = 0.1$. In this experiment, we set $\alpha(1) = 1$ and $\alpha(2) = -1$

. Observe that all of these models under the alternative generate I(0) series, in the sense that their memory decay is exponential.

Figures 6.1 to 6.5 shows realizations of 1000 observations for the above processes, together with a pure fractional process, to illustrate the different ways in which spurious long memory might arise.

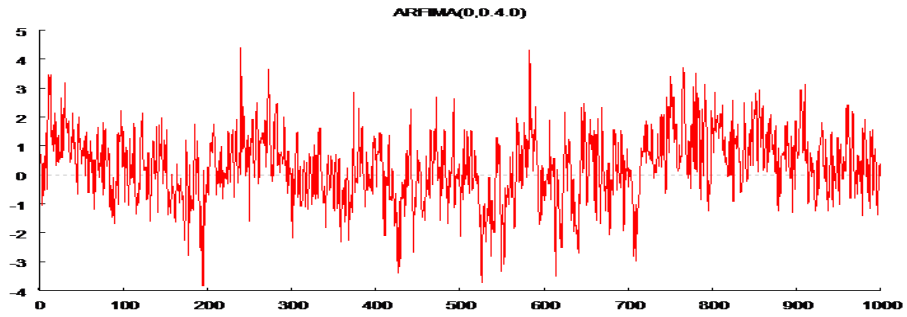


Figure 6.1: ARFIMA (0, 0.4,0)

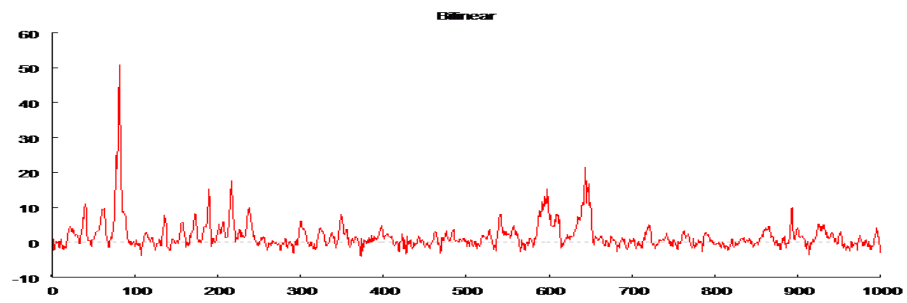


Figure 6.2: Bilinear Model

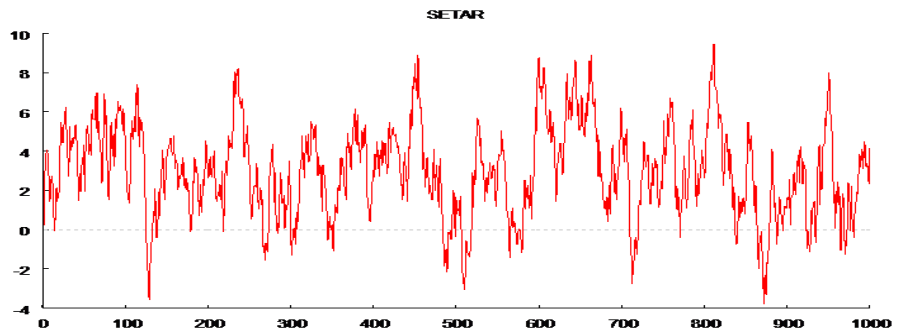


Figure 6.3: SETAR Model

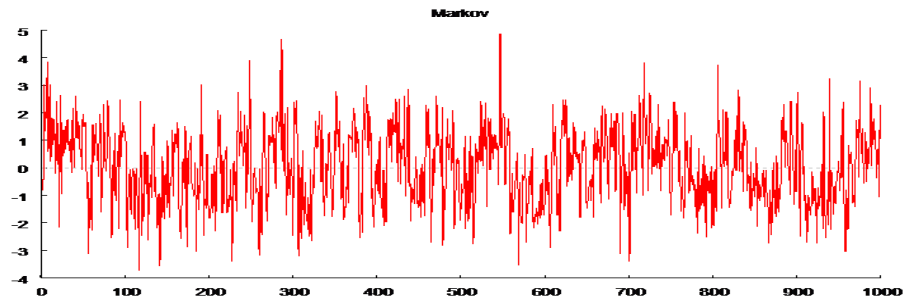


Figure 6.4: Markov Switching

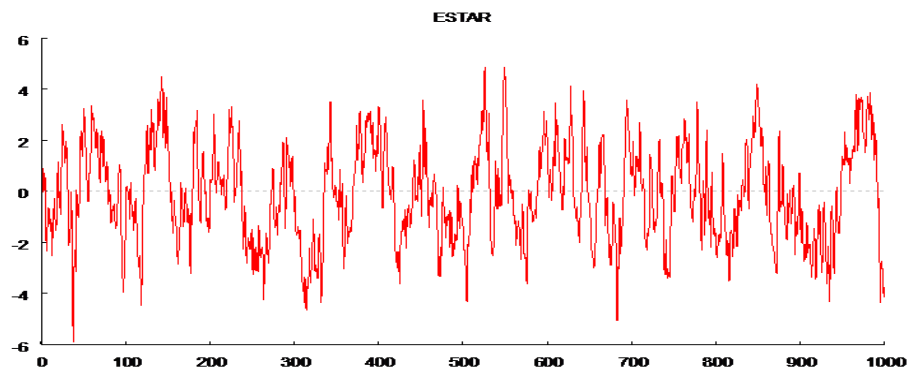


Figure 6.5: ESTAR Model

Table 6.3:

AR(1) Alternatives.

The table illustrates shows the rejection probabilities at the nominal 5% level for autoregressive parameters of 0.8, 0.88 and 0.95. In the case of the alternatives, the figures are not size corrected.

ϕ	T	\hat{d}	M = T ^{0.55}			\hat{d}	M = T ^{0.7}		
			4	8	12		4	8	12
0.8	250	0.32	0.116	0.110	0.086	0.53	0.306	0.283	—
	500	0.23	0.114	0.098	0.078	0.46	0.459	0.402	0.340
	1000	0.15	0.093	0.082	0.062	0.39	0.598	0.506	0.396
	2000	0.09	0.072	0.068	0.057	0.32	0.687	0.554	0.434
0.88	250	0.57	0.047	0.105	0.103	0.70	0.301	0.351	—
	500	0.39	0.050	0.077	0.131	0.63	0.479	0.567	0.496
	1000	0.26	0.170	0.165	0.149	0.55	0.898	0.780	0.704
	2000	0.19	0.149	0.136	0.108	0.48	0.872	0.875	0.856
0.95	250	0.78	0.132	0.158	0.132	0.87	0.158	0.181	—
	500	0.71	0.227	0.240	0.220	0.82	0.326	0.388	0.398
	1000	0.60	0.309	0.363	0.347	0.77	0.582	0.728	0.759
	2000	0.45	0.388	0.440	0.388	0.71	0.833	0.944	0.965

To an extent, the eye is often the best guide to the characteristic appearance of hyperbolic memory.

In Table 6.5, we report some experiments in which the test is treated as the optional second stage of a two-stage procedure. A significance (1-tailed t-test) test is conducted on the full sample log-periodogram estimator. this test is also computed with the bootstrap. If this test results in non-rejection at the 5% level, the second stage test result is discarded and the procedure returns "rejection".

The table shows the percentage overall rejections where the second test is assigned a nominal 5 % significance level. In the left-hand columns, the average estimated d value appears as before. This table also includes the case of the alternative $\phi = 0.7$, which was not

Table 6.4:

Nonlinear Dynamic Models.

The table shows the rejection probabilities for the Bilinear, SETAR, ESTAR and Markov Regime Switching models at the 5 % significance level.

	T	\hat{d}	$M = [T^{0.55}]$			\hat{d}	$M = [T^{0.7}]$		
			$n = 4$	$n = 8$	$n = 12$		$n = 4$	$n = 8$	$n = 12$
Bilinear	250	0.34	0.099	0.073	0.057	0.65	0.449	0.389	–
	500	0.24	0.134	0.111	0.082	0.54	0.596	0.553	0.483
	1000	0.17	0.112	0.107	0.080	0.46	0.735	0.675	0.613
	2000	0.11	0.075	0.070	0.059	0.38	0.814	0.747	0.676
SETAR	250	0.58	0.144	0.168	0.148	0.72	0.256	0.305	–
	500	0.45	0.195	0.200	0.177	0.66	0.464	0.528	0.508
	1000	0.34	0.214	0.218	0.178	0.60	0.706	0.772	0.762
	2000	0.24	0.187	0.187	0.144	0.54	0.897	0.926	0.907
ESTAR	250	0.47	0.140	0.131	0.116	0.67	0.132	0.297	–
	500	0.38	0.163	0.168	0.143	0.60	0.473	0.462	0.490
	1000	0.31	0.167	0.179	0.145	0.53	0.698	0.731	0.691
	2000	0.18	0.137	0.134	0.099	0.46	0.860	0.875	0.842
Markov	250	0.59	0.101	0.242	0.332	0.45	0.126	0.258	–
	500	0.58	0.282	0.439	0.565	0.48	0.281	0.508	0.648
	1000	0.54	0.388	0.539	0.779	0.49	0.542	0.850	0.919
	2000	0.46	0.586	0.852	0.880	0.51	0.871	0.989	0.997

studied previously.

As noted above, this procedure necessarily yields a consistent test, but there is a trade-off of power and size. One main interest is to determine the size characteristics, although given the composite nature of the null, it is difficult to study these except experimentation. The interesting feature of this table, comparing the case of $\phi = 0.8$ with the results in table 6.3, is that the narrow band option looks more attractive (especially assuming better size characteristics) while the performances with the broader bandwidth are similar, thanks to the greater number of spurious first-stage rejections. In the case of $\phi = 0.7$, the effect of spurious rejection is dramatic. The

Table 6.5:

The Two-stage Test procedure, with Skip-rate N=8.

The table illustrates the rejection levels for the autoregressive model when the skip rate of $N = 8$ is adopted.

m	T	Null (FI) $d = 0.4$	Alternatives AR(1)			
			$\phi = 0.7$		$\phi = 0.8$	
			\hat{d}	$n = 8$	\hat{d}	$n = 8$
[$T^{0.55}$]	250	0.307	0.18	0.752	0.39	0.499
	500	0.159	0.12	0.874	0.22	0.599
	1000	0.076	0.06	0.908	0.14	0.678
	2000	0.061	0.04	0.928	0.08	0.786
[$T^{0.7}$]	250	0.087	0.39	0.223	0.54	0.352
	500	0.059	0.31	0.241	0.45	0.467
	1000	0.052	0.24	0.216	0.38	0.515
	2000	0.057	0.19	0.180	0.31	0.558

skip-sampling has little power in this region, and the power actually diminishes as T increases, while the rate of spurious evidently too large to compensate. Eventually the rejection rate must rise again, but evidently with very large samples.

On the basis of the experiments, the best rule of thumb for effectively trading size and power appears to be to use the composite test only in the context of narrow bandwidth estimation, but stick to their broader bandwidth otherwise. There is not a great deal to choose between the rates of skip-sampling, but $N=8$ appears to offer a reasonable of advantages, independent of sample size.

6.9 Application

6.9.1 Individual Stock Prices

In this section, long memory is tested in the volatility of returns for three companies in the IT sector, namely Apple, IBM and Google. The sample size runs from 03/01/2001- 31/12/2010 for Apple and IBM. In the case of Google, the sample size runs from 20/08/2004- 31/12/2010. The data was retrieved from CRSP. Figure 6.6 shows the long memory estimation from the GPH at $M = T^{0.7}$.

Table 6.6:

Application to Individual Absolute Returns.

The table shows an applied case of the test for three individual stock series. The p-value for the ‘fractional’ d is shown in brackets. The ‘Bias test’ column illustrate the p-values from the test of Davidson and Sibbertsen (2008). The skip sampling test was performed on the two bandwidth levels with $N= 4$ and 8.

Company	d	Bias Test	$T^{0.55}$		$T^{0.7}$	
			$n = 4$	$n = 8$	$n = 4$	$n = 8$
Apple	0.141(0)	0.136	0.941	0.729	0.945	0.925
IBM	0.305(0)	0.357	0.382	0.779	0.884	0.949
Google	0.307(0)	0.146	0.578	0.628	0.523	0.397

The results show that there is long memory in log of absolute returns. First of all, this is shown by the statistically significant d . The bias test does not reject the null of long memory in all the cases. Moreover, the skip sample test shows stronger nonrejection of the null hypothesis than the bias test.

6.9.2 Expected Returns

The second application considered is on the expected returns series derived from the previous chapter. We apply the skip sampling test to the expected returns from both models (Kalman Filter and 2 step

Model) for the period 1926-2008. The test is reported illustrated in table 6.7.

Table 6.7:

Application to Expected Returns.

The ‘ d ($T^{.55}$)’ column refers to the Geweke Porter-Hudak estimate for each of the filtered expected returns series. The first two reported tests are the Bias test as in Davidson and Sibbertsen (2008). The last column is the skip sampling test. In all the three cases, the null hypothesis is that the series is a pure long memory process. The corresponding p-values are in brackets.

	d	Bias Test (1)	Bias Test (2)	Skip Sampling Test
State Space AR(1) PD	0.94(0)	1.75(0.04)	1.56(0.06)	0.32(0)
State Space ARFIMA (1,d,0) PD	0.24(0)	-0.78(0.22)	-0.23(0.41)	0.05(0.38)
State Space AR(1) PE	0.52(0)	-0.25(0.39)	-0.03(0.49)	-0.23(0.92)
State Space ARFIMA (1,d,0) PE	0.29(0)	0.62(0.27)	-0.02(0.49)	-0.07(0.61)
2 stage Model 1 (PD)	0.70(0)	1.81(0.03)	0.97(0.16)	0.19(0)
2 stage Model 2 (PD)	0.76(0)	1.28(0.1)	0.42(0.33)	0.17(0)
2 stage Model 3 (PD)	0.26(0)	1.31(0.09)	-0.45(0.32)	-0.01(0.17)
2 stage Model 1 (PE)	0.37(0.18)	2.61(0)	2.56(0)	0.58(0)
2 stage Model 2 (PE)	0.36(0.18)	2.62(0)	2.56(0)	0.58(0)
2 stage Model 3 (PE)	0.44(0.11)	0.76(0.23)	-0.69(0.24)	0.10(0.61)

The semiparametric estimation offers mixed results as to whether expected returns is a fractional process. We analyze the reported ‘ d ’ first. Four models show that there is a possibility of expected returns being a nonstationary process because of an infinite variance ($d > 0.5$). The first two models are the state space models where expected returns were assumed to follow an AR(1). It is highly likely that these processes include both short and long run dependence, which may inflate the semiparametric estimate of d . The same finding occurs for the 2 stage model for constant and AR(1) dividend growth forecast. In such a case, the fit of ARFIMA (1,d,0) appears to be reasonable in modelling the processes. The remaining

six series of expected returns offer a more reasonable d , although the significance of the d parameter from earnings growth forecasts tends to be relatively low.

In order to test whether the expected returns are ‘long memory’, the Bias test (Davidson and Sibbertsen 2006) and the skip sampling test (to be explained more thoroughly in the next chapter) are considered. The Bias test posits using $T/2$ log periodogram points in order to test the null hypothesis of a pure fractional process. According to the test, most of the variables follow a pure fractional process, although it must be admitted that the computed p-value is not far from the rejection zone. Strong rejections can be found on the case of the price-earnings ratio 2 stage models 1 and 2.

In order to consider the coexistence of short run variation with long memory, Bias Test 2 is reported. The Bias test 2 is a variant of the first mentioned bias test but with the bandwidth $M = T^{0.8}$. In this case the null hypothesis is that the process may have an ARFIMA (1, d ,0) structure. The null hypothesis is rejected strongly in the state space model with Price-dividend AR(1) and also the 2 step earnings growth forecast. A simple exercise of fitting an ARFIMA (1, d ,0) to the state space expected returns and the expected returns from earnings growth results in better specification. The skip sampling test was performed with $N = 4$ (the window of each sequential draw), given the small sample size. The results show that the state space expected returns have similar variances (hence pure memory) across subsamples, and hence may be long memory. The expected returns from the present value forecasts are also found to be long memory.

6.10 Conclusion

In this chapter, we have investigated the potential of a test for the null hypothesis of long memory, based on the self similarity property of sequences with the hyperbolic memory decay property. The idea is to compare GPH log periodogram estimators in original and skip-sampled versions of the data set. The property of self-similarity implies that if a new series is created from an original series which contains long memory, the new series should have the same autocorrelation structure as the original series. In this sphere a test may be easily implemented to see whether a series possesses the same property. The aliasing phenomenon, which introduces an estimation bias in the skip samples, poses a problem for the implementation of the test. A bias-corrected estimator permits the construction of a statistic that, although not asymptotically pivotal, allows the implementation of a bootstrap test using the sieve-autoregression method to model short-run dependence.

There is a size-power trade-off involved in the choice of bandwidth for the GPH estimation, and quite large samples prove necessary to yield a decent level of rejection under the alternative cases considered. This is the inevitable consequence of the use of a semi-parametric method to construct the statistic, with correspondingly slow convergence to the asymptote. Nonetheless, the test may prove a useful addition to the arsenal of diagnostic procedures for long memory models, including the bias test of Davidson and Sibbertsen (2009) which compares log γ -periodogram estimates with different bandwidths, and the aggregation test of Ohanissian et. al. (2008).

Monte Carlo simulations show that the test has good size and power properties against linear autoregressive alternatives and also some nonlinear models which are known to yield a statistically significant d when conventional t -ratios are used. The composite test, which is a hybrid of the simple skip sampling and rejection of the

null at the first stage, is found to be better still. The current study may be extended to various cases on how the long memory parameter is computed. Moulines and Soulier (1998) and Robinson (1995) extended ways to compute the long memory parameter. It would be interesting to see how the bootstrapped test may be refined in the context of these estimators. Different asymptotics apply for there is no use of spectral models which is inherent in the GPH estimator. However the principle of the bootstrapped test remains the same.

Two applications of the test was performed. In the first case, long memory was tested in the absolute returns of three stocks, where long memory was evidenced using the skip sampling test and the bias test of Davidson and Sibbertsen (2009). In the second application, long memory was tested in expected returns. Mixed findings were observed. Expected returns from the state space was found to be long memory. Expected returns derived when dividend growth was forecast with the present value variables was also found to exhibit long memory.

Chapter 7

Conclusion

The conclusions of this thesis need to be bridged with both academic and professional work. Trading strategies, in the professional sphere, are mostly viewed as chartists systems which lack economic intuitions. On the other hand, academics thirst in explaining how anomalies may be explained with economic models. This thesis accepts the fact that share prices are excessively volatile (instead of explaining why it is the case), and attempts to make a strategy out of this excess volatility. However, different frequencies and a shorter time span for the holding position is considered. Given figure 1.1, it is obvious that the trading strategy will perform well if the holding position is more than one year. In our case, the trading strategy is implemented with higher frequencies, hinting at the possibility that reversion to the fundamental value may occur quicker.

‘What moves share prices ?’ is a question which fascinates economists. According to economic models, prices are a function of discount rates and expected future payoffs. If dividend data were available at very short frequencies, it would be interesting to see how the present value over time looks. The excess volatility phenomenon involved reversion to the present value after a couple of years when annual data was used.

Chapter two shows that the trading strategy is marginally better than the Buy and Hold strategy at monthly frequencies. However, the success of the strategy depends mostly on the way discount rates and dividend forecasts are computed. Using average realized values of returns as the discount rates ensures that an excess return of 1% over the Buy and Hold, after a 0.5% transaction cost is included. The reliability of the rule was tested using the Sweeney Statistic, Sharpe Ratio and a simulation experiment and most of the test showed positive results. The discount rate needs to be used in conjunction with the autoregressive class of forecasting models, such as the simple AR or the AR with valuation ratios. It is interesting to note that the valuation ratios combined with the autoregressive processes has relatively the same success as the simple AR process.

We also highlight the extreme importance of discount rates. Discount rates which were based on simple statistical computation performed better than those based on fundamentals. The way discount rates are computed changes the statistical distribution of the present value. The cointegration framework of Fama and French shows that in the long run the discount rates may be lower than simple average models, given the current levels of dividend growth, earnings growth and the corresponding yields. When the fundamental measures are used, the stock market is overpriced more often.

The simplicity of the strategy is a matter worth noting. Simple average of dividends and dividend forecast models can prove to be economically profitable. It was found that the type of estimation window does not matter much in terms of the economic profit. The ranking of the discount rate and forecasting models were also robust across linear and logarithmic specifications of dividends. One of the most interesting findings is that the strategy does not work better in the case of the annual data. .

The study is simple and may be extended in different ways. First of all, instead of adopting the strategy only in the case of one stock index, different individual stocks may be held and short according to the present value relationship. Portfolios may be formed on the basis of stocks which are underpriced and overpriced, as in Bulkley and Taylor (1996). Moreover, a simple extension may be to devise a portfolio where the weights to be included in the stock market index and the bond market index is determined as a function of the forecast error between realized dividends and the forecast models. While simple autoregressive models may work well for the dividend generation process of certain stocks, nonlinear and fractional models may be considered.

Chapter three illustrated another model which may be used to derive the discount rates. The expected returns and expected dividend growth rate are derived from the present value. The key assumptions are the time series specifications of both processes, and also the identity between the price dividend ratio, returns and dividend growth in the presence of the present value. The model filters expected returns and dividend growth from the present value. The filtration is based on a vector of optimized parameters from the dataset. The corresponding output is a filtered time series of expected returns and dividend growth. It is found that expected returns is close to being nonstationary, while dividend growth shows no persistence.

The new series of expected returns is used as the discount factor. The difference between the latent variables approach compared to the Cointegration approach of Fama and French(2002) is that in this case the expected returns and dividend growth are jointly modeled with the present value. No such assumption is made in the cointegration approach. The trading strategy was implemented using the

new series of expected returns and expected dividend growth. The trading strategy performed worse than Buy and Hold. Similar to the cointegration approach, the expected returns was much lower than that of the econometric models. It should be stated that in this case, although expected returns and dividend growth are time varying, the present value model is optimized over the whole sample. The parameters are fixed throughout the sample.

Chapter four delved in the predictability of returns from lagged expected returns from Ordinary Least Squares and a Vector Autoregression. The latter ensures that feedback effects are taken into account. Predictability was analyzed both insample and out-of-sample. The results were compared to that of the price dividend ratio. Expected returns is marginally weaker than the price-dividend ratio both in-sample and out of sample. The predictive ability of expected returns appears better in the long run. The success of the price-dividend ratio suggests a possibility of dividend growth containing incremental information over expected returns which helps in predicting returns. The VAR model displayed higher R-squared insample and a lower mean squared error out-of-sample which shows that feedback effects matter.

Chapter five introduced another alternative model for expected returns. Although an autoregressive model appears to be standard model in the finance literature, the AR(1) series may boost up the autoregressive parameter such that expected returns is non-stationary, which is evidenced by some of the stationarity tests. Based on a paper by Granger who proved that aggregated series may be long memory, an alternative specification considered was an ARFIMA(p, d, q). Two models of expected returns was considered. In the first case, expected returns was modeled by an ARFIMA(1,d,0) with the Kalman Filter. In the second case, ex-

pected returns were retrieved from the present value identity.

Two time periods were considered for robustness, and the price earnings ratio was considered as well as the price dividend ratio. The ARFIMA model from the Kalman Filter yields a stationary memory parameter. Moreover, the filtered series are shown to be stationary in three cases. However, the model appears unstable across the different sample sizes and between the two valuation ratios. Three applications were considered for the expected returns: analyzing predictability. Again, it was found that the filtered series were not better than the valuation ratios in terms of insample predictability. Secondly, the expected returns was applied in the trading strategy explored in chapters two and three. It should be noted that the strategies are different in this case, since the model uses a smaller sample size and also uses only annual data. The model parameters are highly dependent on the sample size. Again, the strategy was found to be weaker. lastly, the relationship between expected returns and consumption was investigated. Both the AR and ARFIMA showed impulse response functions which were similar to that of exhibited by Business cycles.

Chapter six is more theoretical in nature. It explains a new bootstrap test of long memory based on aggregation. It uses skip sampled estimators (which induce true long memory) to construct a statistic to determine whether long memory could be spurious. The distribution of estimated d , being pivotal on the sample size and the bandwidth length, requires bias correction because of aliasing problems. The test performs well in large sample sizes. Two applications were considered: The test was used to test for long memory in the absolute returns of three stocks. Most importantly, the test was applied to the expected returns series derived from chapter five. In five cases, the null hypothesis of pure long memory is not rejected. The

null hypothesis is not rejected in the case of the 2 step procedure.

An important limitation is to recognize that asset pricing models rely on assumptions. The assumptions in this thesis were mostly empirical in nature, such as reversion towards fundamentals and the specification of the latent processes. Different categories of agents may hinder reversion. For instance, speculators may drive the price up in the next period, and rational non-speculators may update their information sets without recognizing that it may not reflect the fundamentals. In such a case the present value may be refined with the possibility of a bubble. The fundamental value in itself is moving over time. The study can be refined empirically by assuming different combination of forecast models as in Timmermann (2008) for returns predictability. Various nonlinear specifications of the latent processes, if not computationally demanding, may be used to check strength of the present value relationship.

It is interesting to see how financial variables can exhibit patterns such as long memory. However, theory has little to say about long memory in latent variables. While this thesis makes an attempt in discussing the general empirical modeling of fractionally integrated models in expected returns, a good grounding work is clearly needed, which starts with the aggregation of agent's expectations.

References

- Allen, D., Yang, W., 2004. Do UK. stock prices deviate from fundamentals? *Mathematics and Computers in Simulation* 64 (3-4), 373 – 383.
- Amihud, Y., Hurvich, C. M., 2004. Predictive regressions: A reduced-bias estimation method. *Journal of Financial and Quantitative Analysis* 39 (04), 813–841.
- Ang, A., Bekaert, G., 2007. Stock return predictability: Is it there? *Review of Financial Studies* 20 (3), 651–707.
- Avramov, D., 2002. Stock return predictability and model uncertainty. *Journal of Financial Economics* 64 (3), 423 – 458.
- Avramov, D., 2004. Stock return predictability and asset pricing models. *The Review of Financial Studies* 17 (3), 699–738.
- Baillie, R. T., Bollerslev, T., Mikkelsen, H. O., 1996. Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 74 (1), 3 – 30.
- Baker, M., Wurgler, J., 2000. The equity share in new issues and aggregate stock returns. *The Journal of Finance* 55 (5), 2219–2257.
- Beran, J., 1994. *Statistics for long memory processes*. Chapman and Hall, New York, NY, USA.

- Binsbergern, J. H. V., Koijen, R., 2010. Predictive regressions: A present-value approach. *Journal of Finance* 65 (4), 1439–1471.
- Bossaerts, P., Hillion, P., 1999. Implementing statistical criteria to select return forecasting models: what do we learn? *Rev. Financ. Stud.* 12 (2), 405–428.
- Brandt, M. W., Kang, Q., 2004. On the relationship between the conditional mean and volatility of stock returns: A latent VAR approach. *Journal of Financial Economics* 72 (2), 217 – 257.
- Brockwell, P., Davis, R, A., 1991. *Time Series: theory and methods*. Vol. Second Edition. New York.
- Bulkley, G., Taylor, N., 1996. A cross-section test of the present value model. *Journal of Empirical Finance* 2 (4), 295 – 306.
- Bulkley, G., Tonks, I., 1989. Are UK stock prices excessively volatile? trading rules and variance bound tests. *The Economic Journal* 99 (398), 1083–1098.
- Bulkley, G., Tonks, I., 1992. Trading rules and excess volatility. *The Journal of Financial and Quantitative Analysis* 27 (3), 365–382.
- Bühlmann, P., 1997. Sieve bootstrap for time series. *Bernoulli* 3 (2), pp. 123–148.
- Campbell, J. Y., Sep 1991a. Measuring the persistence of expected returns. NBER Working Papers 3305, National Bureau of Economic Research, Inc.
- Campbell, J. Y., 1991b. A variance decomposition for stock returns. *The Economic Journal* 101 (405), pp. 157–179.
- Campbell, J. Y., 2003. Chapter 13 consumption-based asset pricing. In: *Financial Markets and Asset Pricing*. Vol. 1, Part 2 of *Handbook of the Economics of Finance*. pp. 803 – 887.

- Campbell, J. Y., Shiller, R. J., July 1988. Stock prices, earnings, and expected dividends. *Journal of Finance* 43 (3), 661–76.
- Campbell, J. Y., Yogo, M., 2006. Efficient tests of stock return predictability. *Journal of Financial Economics* 81 (1), 27 – 60.
- Caporale, G. M., Gil-Alana, L. A., 2004. Fractional cointegration and tests of present value models. *Review of Financial Economics* 13 (3), 245 – 258.
- Chambers, M. J., 1998. Long memory and aggregation in macroeconomic time series. *International Economic Review* 39 (4), pp. 1053–1072.
- Chan, N. H., Palma, W., 1998. State space modeling of long-memory processes. *The Annals of Statistics* 26 (2), pp. 719–740.
- Cochrane, J., 2005. *Asset Pricing*. Princeton University Press.
- Cochrane, J., 2008. State space vs VAR. Chicago Booth working paper.
- Cochrane, J., 2010. Discount rates. Chicago Booth Working Paper.
- Cochrane, J. H., 1992. Explaining the variance of price-dividend ratios. *The Review of Financial Studies* 5 (2), 243–280.
- Cochrane, J. H., 2011. Discount rates. NBER Working Papers (16972).
- Conrad, J., Kaul, G., 1988. Time-variation in expected returns. *The Journal of Business* 61 (4), pp. 409–425.
- Crato, N., Ray, B. K., 1996. Model selection and forecasting for long-range dependent processes. *Journal of Forecasting* 15 (2), 107–125.

- Cuthbertson, K., Hyde, S., 2002. Excess volatility and efficiency in french and german stock markets. *Economic Modelling* 19 (3), 399 – 418.
- Dahlhaus, R., 1989. Efficient parameter estimation for self-similar processes. *The Annals of Statistics* 17 (4), pp. 1749–1766.
URL <http://www.jstor.org/stable/2241663>
- Davidson, J., 1994. *Stochastic Limit Theory: An Introduction for Econometricians*. Oxford University Press.
- Davidson, J., Hashimzade, N., 2009. Type I and type II fractional brownian motions: A reconsideration. *Computational Statistics and Data Analysis* 53 (6), 2089 – 2106, the Fourth Special Issue on Computational Econometrics.
- Davidson, J., Sibbertsen, P., 2009. Tests of bias in log-periodogram regression. *Economics Letters* 102 (2), 83 – 86.
- Delgado, M. A., Velasco, C., 2005. Sign tests for long-memory time series. *Journal of Econometrics* 128 (2), 215 – 251.
- Diba, B. T., Grossman, H. I., 1988. Explosive rational bubbles in stock prices? *The American Economic Review* 78 (3), pp. 520–530.
- Diebold, F. X., Mariano, R. S., 1995. Comparing predictive accuracy. *Journal of Business and Economic Statistics* 13 (3), 253–263.
- Dolado, J. J., Gonzalo, J., Mayoral, L., 2002. A fractional Dickey - Fuller test for unit roots. *Econometrica* 70 (5), 1963–2006.
- Dufour, J.-M., 1997. Some impossibility theorems in econometrics with applications to structural and dynamic models. *Econometrica*.

- Fama, E. F., French, K. R., April 1988. Permanent and temporary components of stock prices. *Journal of Political Economy* 96 (2), 246–73.
- Fama, E. F., French, K. R., 2002. The equity premium. *The Journal of Finance* 57 (2), 637–659.
- Faust, J., 1996. Near observational equivalence and theoretical size problems with unit root tests. *Econometric Theory* 12 (04), 724–731.
- Faust, J., 1999. Conventional confidence intervals for points on spectrum have confidence level zero. *Econometrica* 67 (3), 629–637.
- Ferson W.E, Sarkissian, S., Simin, T., August 2003. Spurious regressions in financial economics? *The Journal of Finance* 58.
- Gallant, A., White, H., 1988. A unified theory of estimation and inference for nonlinear dynamic models. B.Blackwell.
- Geweke, J., Porter-Hudak, S., 1983. The estimation and application of long memory in time series models. *Journal of Time Series Analysis* 4 (4), 221–238.
- Giraitis, L., Kokoszka, P., Leipus, R., Teyssière, G., 2003. Rescaled variance and related tests for long memory in volatility and levels. *Journal of Econometrics* 112 (2), 265 – 294.
- Goyal, A., Welch, I., 2003. Predicting the equity premium with dividend ratios. *Management Science* 49 (5), 639–654.
- Goyal, A., Welch, I., 2007. A Comprehensive Look at The Empirical Performance of Equity Premium Prediction. *Rev. Financ. Stud.*
- Granger, C. W. J., Joyeux, R., 1980. An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis* 1 (1), 15–29.

- Hannan, E. J., Deistler, M., 1987. The statistical theory of linear systems. John Wiley & Sons, Inc., New York, NY, USA.
- Hansen, P., 2001. An unbiased and powerful test for superior predictive ability. Working paper, Department of Economics, Brown University.
- Hassler, U., June 2011. Estimation of fractional integration under temporal aggregation. *Journal of Econometrics* 162 (2), 240–247.
- Horowitz, J. L., 2001. The bootstrap. Vol. 5 of *Handbook of Econometrics*. Elsevier, pp. 3159 – 3228.
- Hosking, J. R. M., 1981. Fractional differencing. *Biometrika* 68 (1), 165–176.
- Hurvich, C. M., Deo, R., Brodsky, J., 1998. The mean squared error of Geweke and Porter-Hudak’s estimator of the memory parameter of a long-memory time series. *Journal of Time Series Analysis* 19 (1), 19–46.
- Inoue, A., Kilian, L., Dec. 2002. In-sample or out-of-sample tests of predictability: Which one should we use? Tech. Rep. 3671, C.E.P.R. Discussion Papers.
- Kanas, A., 2005. Nonlinearity in the stock price-dividend relation. *Journal of International Money and Finance* 24 (4), 583 – 606.
- Kim, C. S., Phillips, P. C., Oct. 2006. Log periodogram regression: The nonstationary case. Cowles Foundation Discussion Papers 1587, Cowles Foundation for Research in Economics, Yale University.
- Lamont, O., 1998. Earnings and expected returns. *The Journal of Finance* 53 (5), 1563–1587.

- LeRoy, S. F., Porter, R. D., 1981. The present-value relation: Tests based on implied variance bounds. *Econometrica* 49 (3), pp. 555–574.
- Lettau, M., Ludvigson, S., 2001. Consumption, aggregate wealth, and expected stock returns. *The Journal of Finance* 56 (3), 815–849.
- Lettau, M., Van Nieuwerburgh, S., 2008. Reconciling the Return Predictability Evidence. *Rev. Financ. Stud.* 21 (4), 1607–1652.
- Lewellen, J., 2004. Predicting returns with financial ratios. *Journal of Financial Economics* 74 (2), 209 – 235.
- Ljungqvist, L., Sargent, T., 2004. *Recursive Macroeconomic Theory*. MIT Press.
- Lo, A., 1991. Long term memory in stock market prices. *Econometrica* 59, 1279–1313.
- Lobato, I. N., Robinson, P. M., 1998. A nonparametric test for $i(0)$. *The Review of Economic Studies* 65 (3), 475–495.
- Lobato, I. N., Velasco, C., 2000. Long memory in stock market trading volume. *Journal of Business and Economic Statistics* 18 (4), pp. 410–427.
- Lustig, H. N., Nieuwerburgh, S. G. V., 2005. Housing collateral, consumption insurance, and risk premia: An empirical perspective. *Journal of Finance* 60 (3), 1167–1219.
- Mankiw, N. G., Romer, D., Shapiro, M. D., 1985. An unbiased reexamination of stock market volatility. *The Journal of Finance* 40 (3), 677–687.
- Menzly, L., Santos, T., Veronesi, P., 2004. Understanding predictability. *Journal of Political Economy* 112 (1), 1–47.

- Mills, T. C., 1993. Testing the present value model of equity prices for the uk stock market. *Journal of Business Finance and Accounting* 20 (6), 803 – 813.
- Moulines, E., Soulier, P., 1999. Broadband log-periodogram regression of time series with long-range dependence. *The Annals of Statistics* 27 (4), pp. 1415–1439.
- Ohanissian, A., Russell, J. R., Tsay, R. S., 2008. True or spurious long memory? A new test. *Journal of Business and Economic Statistics* 26, 161–175.
- Pastor, L., Stambaugh, R. F., Jan. 2007. Predictive systems: Living with imperfect predictors. NBER Working Papers 12814, National Bureau of Economic Research, Inc.
- Pesaran, H., Pettenuzzo, D., Timmermann, A., 2007. Learning, structural instability, and present value calculations. *Econometric Reviews* 26 (2-4), 253 – 288.
- Pesaran, M. H., Timmermann, A., 2002. Market timing and return prediction under model instability. *Journal of Empirical Finance* 9 (5), 495 – 510.
- Piazzesi, M., Schneider, M., Tuzel, S., 2007. Housing, consumption and asset pricing. *Journal of Financial Economics* 83 (3), 531 – 569.
- Polk, C., Thompson, S., Vuolteenaho, T., 2006. Cross-sectional forecasts of the equity premium. *Journal of Financial Economics* 81 (1), 101 – 141.
- Poskitt, D., 2007. Autoregressive approximation in nonstandard situations: the fractionally integrated and non-invertible cases. *Annals of the Institute of Statistical Mathematics* 59, 697–725.

- Poterba, J. M., Summers, L. H., 1988. Mean reversion in stock prices : Evidence and implications. *Journal of Financial Economics* 22 (1), 27 – 59.
- Pötscher, B. M., 2002. Lower risk bounds and properties of confidence sets for ill-posed estimation problems with applications to spectral density and persistence estimation, unit roots, and estimation of long memory parameters. *Econometrica* 70 (3), 1035–1065.
- Rangvid, J., 2006. Output and expected returns. *Journal of Financial Economics* 81 (3), 595 – 624.
- Ray, B. K., 1993. Modeling long memory processes for optimal long-range prediction. *Journal of Time Series Analysis* 14 (5), 511–525.
- Rey, D., October 2005. Market timing and model uncertainty: An exploratory study for the swiss stock market. *Financial Markets and Portfolio Management* 19 (3), 239–260.
- Robinson, P. M., 1995. Log-periodogram regression of time series with long range dependence. *The Annals of Statistics* 23 (3), pp. 1048–1072.
- Rytchkov, O., 2006. Filtering out expected dividends and expected returns. MIT Working paper.
- Shiller, R. J., 1981. Do stock prices move too much to be justified by subsequent changes in dividends? *The American Economic Review* 71 (3).
- Shiller, R. J., Beltratti, A. E., Apr. 1993. Stock prices and bond yields: Can their comovements be explained in terms of present value models?

- Souza, L. R., 2005. A note on chambers's long memory and aggregation in macroeconomic time series. *International Economic Review* 46 (3), 1059–1062.
- Souza, L. R., Smith, J., 2002. Bias in the memory parameter for different sampling rates. *International Journal of Forecasting* 18 (2), 299 – 313.
- Souza, L. R., Smith, J., 2004. Effects of temporal aggregation on estimates and forecasts of fractionally integrated processes: a monte-carlo study. *International Journal of Forecasting* 20 (3), 487 – 502.
- Stambaugh, R. F., 1999. Predictive regressions. *Journal of Financial Economics* 54 (3), 375 – 421.
- Strauss, J., Yigit, T., 2001. Present value model, heteroscedasticity and parameter stability tests. *Economics Letters* 73 (3), 375 – 378.
- Sweeney, R. J., 1986. Beating the foreign exchange market. *The Journal of Finance* 41 (1), 163–182.
- Taylor, N., 2012. A formula for the economic value of return predictability. *The European Journal of Finance* 0 (0), 1–17.
- Timmermann, A., October 1996. Excess volatility and predictability of stock prices in autoregressive dividend models with learning. *Review of Economic Studies* 63 (4), 523–57.
- Timmermann, A., 2008. Elusive return predictability. *International Journal of Forecasting* 24 (1), 1–18.
- Timmermann, A., Granger, C. W. J., 2004. Efficient market hypothesis and forecasting. *International Journal of Forecasting* 20 (1), 15 – 27.

- Timmermann, A. G., 1993a. How learning in financial markets generates excess volatility and predictability in stock prices. *The Quarterly Journal of Economics* 108 (4), 1135–1145.
- Timmermann, A. G., November 1993b. How learning in financial markets generates excess volatility and predictability in stock prices. *The Quarterly Journal of Economics* 108 (4), 1135–45.
- Valkanov, R., 2003. Long-horizon regressions: theoretical results and applications. *Journal of Financial Economics* 68 (2), 201 – 232.
- Van Binsbergen, J. H., Koijen, R. S., 2009. Predictive Regressions: A Present-Value Approach. SSRN eLibrary.
- Velasco, C., 1999. Non-stationary log-periodogram regression. *Journal of Econometrics* 91 (2), 325 – 371.

Appendix A

Essays in Asset Pricing

The appendix is divided according to notes to the different chapters. They illustrate statistics, proofs and algebraic derivations which may be of interest.

A.1 Appendix 1

A.1.1 Holding Return

Table A.1 shows the percentage average return over different months. For instance, for the 36 month period, investors receive 37.5 % over their initial investment in year zero if forecasting model 1a (Recursive) and Model A of the discount rate is used. In all the cases, discount factor A and B along with forecasting model 1 and 2. The rate of return between the best and worst strategy tends to grow larger with time. It simply shows that the trading strategy performs well in the long run.

Table A.1:

Cumulated Returns (in percentage) over 5 years of Holding.

The table shows the rate of returns for the strategy over the different holding horizons. The results for reported for both the discount rates and the forecast models.

Model	24 Months				36 Months				48 Months				60 Months			
	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D
1a Recursive	23.6	21	14.4	17.7	37.5	33.2	22.2	27.5	53.6	46.9	30.7	38.5	71	61.4	39.6	50
1a Rolling	23.7	20.8	14.4	17.7	37.7	32.9	22.2	27.5	53.8	46.5	30.7	38.5	71.3	60.9	39.6	50
2a Recursive	23.7	20.8	14.4	17.7	37.7	32.9	22.2	27.5	53.8	46.5	30.7	38.5	71.3	60.9	39.6	50
2a Rolling	23.5	21.1	14.1	17.5	37.4	33.4	21.7	27.1	53.2	47.2	29.9	38.0	70.4	61.9	38.6	49.3
3a Recursive	15.5	15.4	12.1	15.5	24.2	23.9	18.6	24	33.6	33.0	25.5	33.3	43.9	43	32.6	42.9
3a Rolling	17.4	17.1	14.0	15.5	27.4	26.8	21.7	24.2	38.1	37.2	29.9	33.5	49.7	48.3	38.6	43.1
1b Recursive	23.4	21.6	14.4	17.2	37.2	34.3	22.2	26.8	52.9	48.4	30.6	37.5	70.2	63.6	39.5	48.6
1b Rolling	23.4	21.6	14.4	17.2	37.2	34.3	22.2	26.8	52.9	48.4	30.6	37.5	70.2	63.6	39.5	48.6
2b Recursive	23.4	21.1	14.1	17.2	37.3	33.4	21.7	26.8	53.2	47.1	29.9	37.5	70.5	61.9	38.5	48.6
2b Rolling	23.4	21.1	14.1	17.2	37.3	33.4	21.7	26.8	53.2	47.2	29.9	37.5	70.5	62	38.5	48.6
3b Recursive	15.8	14.9	12.8	15.5	24.6	23.1	19.8	24.0	34	31.9	27.1	33.3	44.3	41.6	34.8	42.9
3b Rolling	18.3	18	14.2	15.8	28.6	28.1	22	24.6	39.7	38.9	30.3	34.2	51.6	50.4	39	43.9
Buy and Hold	15.1				23.6				32.6				42.2			

A.1.2 Graphical Plots of Accumulated Wealth

The graphs show the evolution of wealth from January 1901 until December 2007. They illustrate the wealth from Buy and Hold (W_{rm}), and the wealth under the different forecasting models and discount rates.

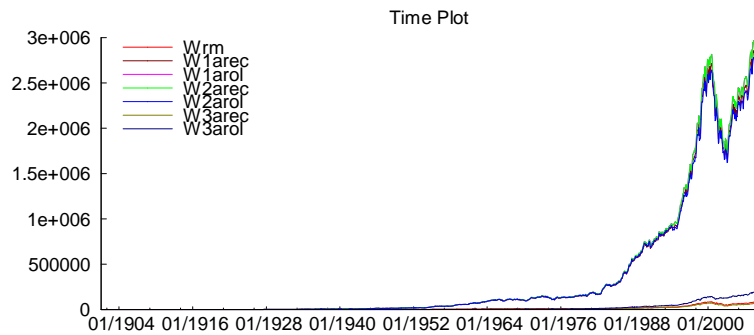


Figure A.1: Cumulated Wealth from January 1901 to December 2007 for Discount Rate A and Forecast Model A .

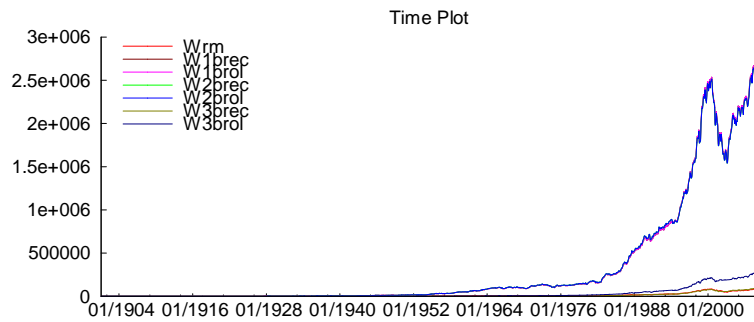


Figure A.2: Cumulated Wealth from January 1901 to December 2007 for Discount Rate A and Forecast Model B.

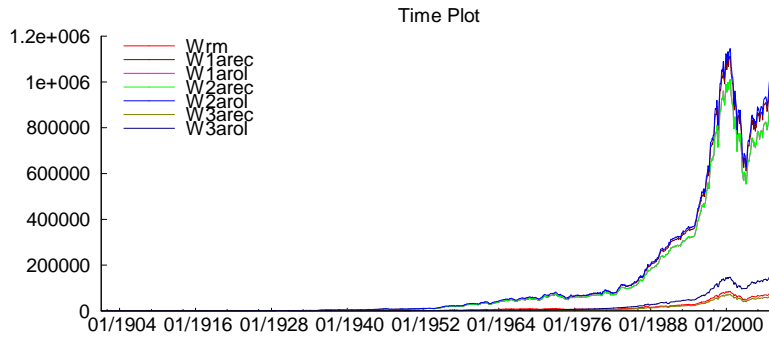


Figure A.3: Cumulated Wealth from January 1901 to December 2007 for Discount Rate B and Forecast Model A.

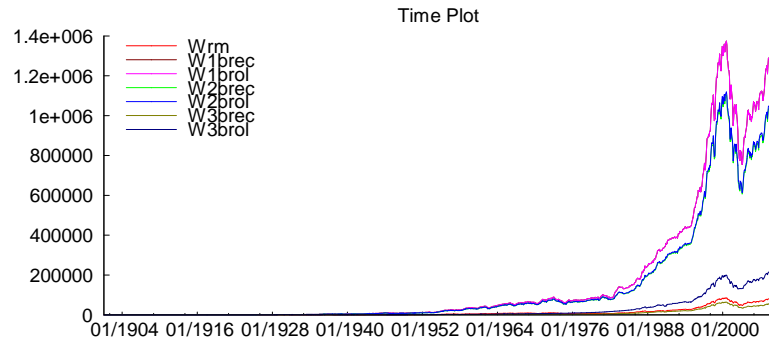


Figure A.4: Cumulated Wealth from January 1901 to December 2007 for Discount Rate B and forecast model B.

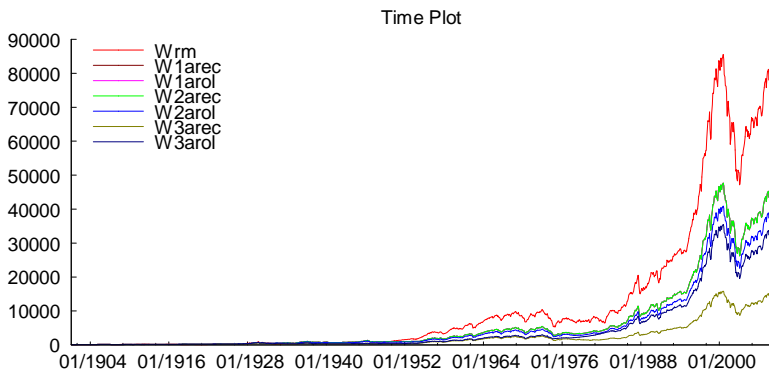


Figure A.5: Cumulated Wealth from January 1901 to December 2007 for Discount Rate C and Forecast Model A.

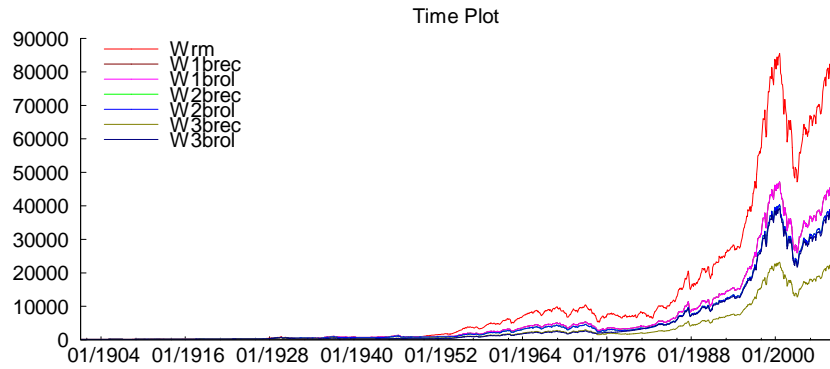


Figure A.6: Cumulated Wealth from January 1901 to December 2007 for Discount Rate C and Forecast Model B

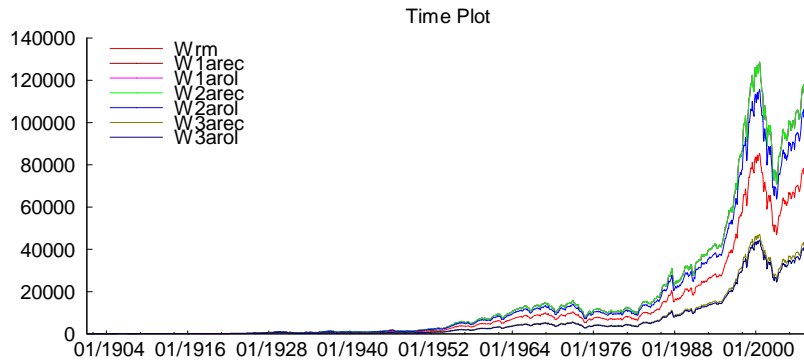


Figure A.7: Cumulated Wealth from January 1901 to December 2007 for Discount Rate D and Forecast Models A.

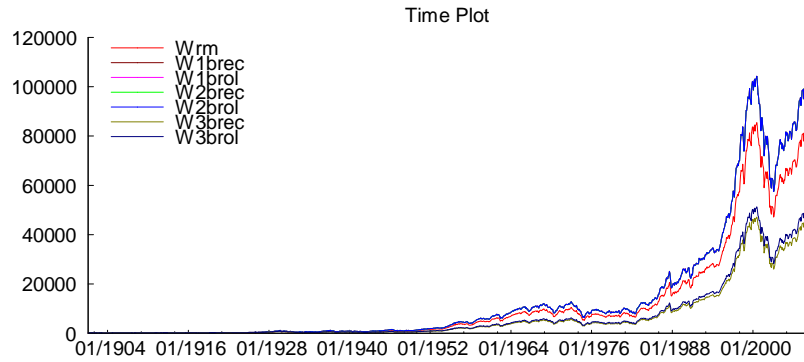


Figure A.8: Cumulated Wealth from January 1901 to December 2007 for Discount Rate D and Forecast Model B.

The graphs depict high performance from models 1 and 2 for both the logarithmic and levels for denominators A and B. On the other hand, all models tend to perform rather poorly for denominators C and D, although the hierarchy of the best forecasting model is still 1 and 2. Denominators C and D exhibit a lower variance across the different models in terms of accumulated wealth. The graphs show that the strategy tends to pick up during the great depression and the financial crash of 1939-42. The trading strategy appears to fare badly during the dot.com bubble burst.

A.1.3 Test of Correlated Means

Table A.2:

Tests of Correlated Means.

The table shows the number of times, the strategy beats, equals, or is outperformed by Buy and Hold.

	12 Months	24 Months	36 Months	48 Months	60 Months
$R_{tr} > R_m$	16	19	22	21	21
$R_{tr} < R_m$	20	20	20	20	20
$R_{tr} = R_m$	12	9	6	7	7

Table A.2 summarizes the results from table A.2 and defines the number of times each of the three different scenarios was witnessed. $R_{tr} < R_m$ is the number of times that the accumulated return is higher than the trading rule. This proportion is high due to the adoption of denominator C and D. It can be seen that as the horizon increases, there are fewer instances when $R_{tr} = R_m$. During the one year period, it is likely that for a whole year, the asset is kept in equity.

Table A.3:

Z-score for Individual models.

The table shows the computed $Z_{l,d}(k)$ statistic for different k , l and d , where k is the horizon, l is the forecasting model and d is the discount factor. $Z_{l,d}(k)$ follows a normal distribution. The null hypothesis is that both the market and trading rule return are equal in each period. A negative statistic implies that the trading rule yields a higher market return.

Model	1a Rec	1a Rol	2a Rec	2a Rol	3a Rec	3a Rol	1b Rec	1b Rol	2b Rec	2b Rol	3b Rec	3b Rol
1 year :A	-10.29	-10.43	-10.43	-10.27	2.81	0.18	-9.99	-9.99	-10.14	-10.14	2.22	-1.06
B	-5.96	-5.56	-5.56	-6.05	2.76	0.58	-6.87	-6.87	-5.85	-5.92	3.22	-0.59
C	4.12	4.12	4.12	4.33	6.76	4.52	4.16	4.16	4.36	4.36	5.75	4.20
D	-1.08	-1.08	-1.08	-0.25	3.79	3.97	0.53	0.53	0.53	0.53	3.79	3.59
2 year :A	-14.33	-14.51	-14.51	-14.20	3.41	-0.25	-14.13	-14.13	-14.39	-14.39	2.75	-1.90
B	-8.45	-7.97	-7.97	-8.71	3.63	0.34	-9.76	-9.76	-8.49	-8.56	4.43	-1.31
C	6.03	6.03	6.03	6.41	10.32	6.44	6.07	6.07	6.46	6.46	8.72	5.90
D	-3.00	-3.00	-3.00	-1.36	5.57	5.36	0.08	0.08	0.09	0.08	5.57	4.84
3 year :A	-18.67	-18.92	-18.92	-18.24	3.57	-1.00	-18.60	-18.60	-18.99	-18.99	2.97	-2.84
B	-11.13	-10.65	-10.65	-11.69	4.03	-0.16	-12.94	-12.94	-11.43	-11.49	4.90	-2.07
C	7.26	7.26	7.26	7.86	12.36	7.59	7.31	7.31	7.91	7.91	10.35	6.94
D	-4.53	-4.53	-4.53	-2.51	7.07	6.56	-0.45	-0.45	-0.44	-0.45	7.07	5.80
4 year :A	-21.84	-22.08	-22.08	-21.03	4.07	-0.75	-21.87	-21.87	-22.33	-22.33	3.50	-2.60
B	-12.19	-11.71	-11.71	-12.92	4.67	0.21	-14.46	-14.46	-12.77	-12.84	5.49	-1.66
C	8.14	8.14	8.14	8.73	13.31	8.41	8.19	8.19	8.78	8.78	11.27	7.81
D	-4.45	-4.45	-4.45	-2.23	7.83	7.48	-0.16	-0.16	-0.15	-0.16	7.83	6.80
5 year :A	-23.41	-23.63	-23.63	-22.58	3.87	-1.20	-24.16	-24.16	-24.28	-24.28	3.44	-3.13
B	-13.66	-13.15	-13.15	-14.57	4.73	0.02	-16.36	-16.36	-14.49	-14.55	5.74	-1.90
C	8.40	8.40	8.40	9.21	14.62	9.25	8.45	8.45	9.26	9.26	12.43	8.56
D	-5.97	-5.97	-5.97	-3.22	8.96	8.56	-0.85	-0.85	-0.84	-0.85	8.96	7.89

The Z-score is higher as the horizon expands. Forecast model 3a and 3b with recursive windows is positive (Buy and Hold is better than the trading strategy) for all denominators and horizons.

A.1.4 Sweeney X Statistic

One of the measures to account for the fact that wealth is held in both the stock market and in the bond market during the period of interest is given by the Sweeney(1989) statistic. The X statistic is given by:

$$X = R^{tr} - (1 - f)R^{BH}$$

205

$$\sigma_x = \sigma[f(1 - f)/N]^{\frac{1}{2}}$$

The Sweeney's statistic is computed in this case assuming a worst case scenario of 5 % in terms of transaction costs. In this case, a 5 percentage point is subtracted from the trading rule returns. The Sweeney statistic is negative in all cases.

Table A.4:

Sweeney Statistic assuming 5 % Transaction cost.

The Sweeney Statistic is computed assuming that there is a cost of 5% when the agent switches the asset. The Sweeney statistic is computed with the adjusted returns.

	1aRec	1aRol	2aRec	2aRol	3aRec	3aRol	1bRec	1bRol	2bRec	2bRol	3bRec	3bRol
A	-2.603	-2.533	-2.533	-2.627	-6.781	-3.932	-2.669	-2.669	-2.705	-2.718	-5.704	-3.451
B	-4.155	-4.289	-4.289	-4.091	-6.741	-4.154	-3.757	-3.757	-4.146	-4.142	-6.373	-3.714
C	-9.323	-9.301	-9.301	-9.560	-9.110	-6.334	-9.275	-9.275	-9.581	-9.604	-7.834	-6.204
D	-7.158	-7.138	-7.138	-7.315	-7.165	-5.626	-7.448	-7.448	-7.488	-7.507	-6.435	-5.391

Table A.5:

Sweeney Statistic assuming 1 % Transaction Cost.

The Sweeney's statistic is computed assuming a monthly rate of 1 % in terms of transaction costs. The table shows that discount rates A and B are still profitable for the successful models. Discount D is marginally profitable. Discount C is negative in all cases.

	1aRec	1aRol	2aRec	2aRol	3aRec	3aRol	1bRec	1bRol	2bRec	2bRol	3bRec	3bRol
A	5.466	5.525	5.525	5.525	0.487	2.810	5.367	5.367	5.354	5.351	1.287	3.293
B	3.914	3.769	3.769	3.769	0.528	2.588	4.280	4.280	3.912	3.927	0.618	3.030
C	-1.254	-1.242	-1.242	-1.242	-1.842	0.408	-1.238	-1.238	-1.523	-1.534	-0.843	0.540
D	0.912	0.920	0.920	0.744	0.104	1.116	0.589	0.589	0.571	0.563	0.556	1.353

Table A.6:

Sweeney Statistic assuming 2.5 % for Transaction cost.

The table shows that discount rates C and D are not profitable at all should the transaction cost be 2.5 %. In this case, only discount rate A and the ARMAX and AR(p) model seem to work.

	1a Rec	1a Rol	2a Rec	2a Rol	3a Rec	3a Rol	1b Rec	1b Rol	2b Rec	2b Rol	3b Rec	3b Rol
A	2.440	2.503	2.503	2.409	-2.238	0.282	2.354	2.354	2.332	2.325	-1.335	0.764
B	0.888	0.747	0.747	0.945	-2.198	0.060	1.266	1.266	0.890	0.901	-2.004	0.501
C	-4.280	-4.264	-4.264	-4.523	-4.568	-2.121	-4.252	-4.252	-4.544	-4.560	-3.465	-1.989
D	2.114	-2.102	-2.102	-2.278	-2.622	-1.412	-2.425	-2.425	-2.451	-2.463	-2.066	-1.176

A.1.5 Switching

Tables A.7 and A.8 respectively show the number of months equity is held and the number of times the strategy postulate switching.

Table A.7:

Equity Holding Periods.

Table A.7 illustrates the number of times that the rule postulates going long on the market for the different forecast models. It is interesting to note that there is no considerable difference between discount rate C and the best performing discount rates A and B. However discount rate D advises going long more often, without better success. In terms of the individual models, worst performing model 3b postulate going long less often.

	A	B	C	D
1a Recursive	883	857	884	1250
1a Rolling	881	858	884	1250
2a Recursive	881	858	884	1250
2a Rolling	884	858	883	1249
3a Recursive	882	881	874	1167
3a Rolling	766	774	863	1179
1b Recursive	885	863	886	1248
1b Rolling	885	863	886	1248
2b Recursive	890	860	884	1247
2b Rolling	890	859	884	1248
3b Recursive	817	823	847	1167
3b Rolling	760	774	858	1190

Table A.8:

Number of Switches

Table A.8 shows the number of times that the rule postulates a ‘switch’ is performed. Given the previous results on terminal wealth, it may be noted that the models which has the highest accumulated wealth has more switches. The worst discount rate model has the few switches although the asset is held in equity as often as denominators A and B. Discount rate A with recursive and rolling forecasts of model 1b suggests more switches marginally.

	A	B	C	D
1a Recursive	420	402	42	10
1a Rolling	418	404	42	10
2a Recursive	418	404	42	10
2a Rolling	416	394	42	10
3a Recursive	92	78	40	10
3a Rolling	54	48	32	12
1b Recursive	422	392	42	10
1b Rolling	422	392	42	10
2b Recursive	412	392	42	12
2b Rolling	412	392	42	10
3b Recursive	102	82	34	10
3b Rolling	50	42	30	16

A.1.6 Tests on Forecasting Accuracy

Tests of normality on the forecasting errors are reported in table A.9.

Table A.9:

Tests of Normality on Forecast Error.

The table shows the statistics from the Anderson-Darling, Kolgomorov-Smirnov and Doornik and Hansen test.

	AD test	Kolgomorov Smirnov	Doornik Hansen test
1a Recursive	20.82	0.09	690.13
1a Rolling	20.92	0.09	688.13
2a Recursive	20.92	0.09	688.13
2a Rolling	20.71	0.09	692.11
3a Recursive	6.08	0.05	3.73
3a Rolling	14.55	0.09	87.69
1b Recursive	12.63	0.07	456.59
1b Rolling	12.63	0.07	456.59
2b Recursive	13.36	0.06	469.29
2b Rolling	13.97	0.07	475.81
3b Recursive	3.88	0.04	4.13
3b Rolling	14.11	0.09	113.27

Table A.9 shows tests on the normality of the forecast errors using the Anderson-Darling (Stephens(1974)), Kolgomorov-Smirnov (Kolgomorov 1933) and Doornik and Hansen (Doornik and Hansen (1994)) procedures. The Anderson-Darling and the Kolgomorov-Smirnov are nonparametric tests for comparing two distributions based on goodness of fit tests for normality. They are based on the on the maximum distance between the empirical distribution and the normal distribution. The Anderson Darling test turns out to be more sensitive to deviations in the tails of the distribution than the Kolmogorov- Smirnov test. The null hypothesis is that the forecast errors come from the normal distribution. The Doornik Hansen Omnibus test of normality based on Shenton and Bowman (1977) is a test based on the skewness and kurtosis of the distribution. They proved that their proposed test has the best size and power properties of the tests considered, even better than the Anderson- Darling and Kolgomorov-Smirnov Test. Most of the tests show that the

forecast errors are not normal. However, it may be seen that recursive windows have a higher level of normality than rolling windows. This may be due to large sample property that recursive windows possess.

Hansen Test of Superior Predictability

Table A.10:

Forecasting Ability test using Hansen's Superior Ability test.

	Model		Sample Loss		T-Statistic	
	MAE	MSE	MAE	MSE	MAE	MSE
Most Significant	2a Rol	2a Rol	0.065	0.009	-26.57	-10.6
Best	2a Rec	2a Rec	0.064	0.009	-26.6	-10.68
Model 25 %	2a Rol	2a Rol	0.065	0.009	-26.57	-10.6
Median	1b Rec	1b Rec	0.071	0.01	-29.4	-11.9
Model 75 %	3b Redc	3a Rec	1.758	4.59	-27.87	-19.51
Worst	3b Rol	3a Rol	2.967	5.84	-32.92	-13.98

The Hansen (2001) test of superior predictive ability requires the definition of a loss function $L(D_t, \widehat{D}_t)$ where D_t and \widehat{D}_t are the realized and predicted values respectively. In the case of the k models we are comparing, we have $2k$ predicted values \widehat{D}_t and hence loss functions accounting for both rolling and recursive windows. The performance of a particular model is compared to the performance of a benchmark model based on a particular loss function. This is given by (A.1).

$$X_k(t) = L(D_t, \widehat{D}_{0t}) - L(D_t, \widehat{D}_{kt}) \quad (\text{A.1})$$

k refers to the specific model we are assessing. Hence $K = \text{M1a...M3b}$
 $t = 1, \dots, n$

The Loss function can take any of the above named forecast performance assessment criteria like the Mean Squared Error(MSE),

Mean Absolute Error (MAE), Mean Error (ME) and Root Mean Squared Error (RMSE). In this study, we use the two most common statistics in the forecasting literature, namely the MSE and MAE.

The null hypothesis for the test of superior ability is that the benchmark model (\widehat{D}_{0t}) is the best model. This is defined as :

$$H_0 : \mu_k = E(X_k(t)) \leq 0,$$

or

$$H_0 : \mu \leq 0, \text{ in vector equivalent}$$

defining the standardized maximum t-statistic:

$$T_n^{sm} = \max_k \frac{n^{\frac{1}{2}} \overline{X}_k}{\widehat{\sigma}_k} \quad (\text{A.2})$$

where

$$\overline{X}_k = \frac{1}{n} \sum_{i=1}^n X_k(t) \quad (\text{A.3})$$

and

$$\widehat{\sigma}_k^2 = \text{var}(n)^{\frac{1}{2}} \overline{X}_k \quad (\text{A.4})$$

which is estimated by a bootstrap method.

The test of Superior Predictive Accuracy shows that the best model is Model 2a (recursive window). Irrespective of the loss function adopted the ranking of the models are similar except for the worse models where the Mean Squared Error loss function states that the worst models are model 1b (the logarithmic counterpart of the real model with dividends). The result that the model 2a is the best confirms the earlier findings of end point of the recursive mean squared error (or the unconditional mean squared error). Adding new macroeconomic variables does not improve the forecast of dividends. However, it should be said that the ARMAX and AR(p) have little difference in their forecasting power.

A.1.7 Summary Statistics

Table A.11:

Descriptive Statistics for Dividends and Log Dividends.

The table illustrates the descriptive statistics of dividends and the logarithm of dividends. Tests of nonstationarity are provided by the Augmented Dickey Fuller test (ADF) and Phillips-Perron test (PP). The results show that the dividends is nonstationary in levels, but may be stationary after detrending. The results show that both dividends and log of dividends are not normally distributed. Dividends and its logarithm tend to harbour a unit root.

	Dividends			Log Dividends		
	Series	Detrended	Differenced	Series	Detrended	Differenced
Mean	12.83	12.83	0.02	2.48	2.48	0.00
Std. Dev	4.66	2.27	0.13	0.37	0.20	0.01
Skewness	0.46	0.11	-0.48	-0.13	-0.59	-0.94
Kurtosis	2.50	2.71	7.12	2.00	2.71	10.25
Jarque-Bera	57.78	6.88	959.95	56.95	78.63	2996.31
ADF test	-0.70	-1.27	-12.43	-0.86	-2.79	-12.12
PP test	1.04	-1.80	-18.34	-0.98	-2.89	-19.16
KPSS test	1.14	0.39	0.20	0.08	0.29	0.06
Lo's Test	1.62	2.02	0.85	0.44	2.10	0.75
Robinson's d	0.49	0.50	0.34	0.50	0.50	0.31

A.1.8 Simulation Results

Table A.12:

Simulation Experiment.

The table shows the number of times (out of 160 trials) that the strategy beats the Buy and Hold for the different discount rates and forecast models.

Model	12 Months				24 Months				36 Months				48 Months				60 Months			
	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D
1a Recursive	67	55	23	5	93	79	34	12	104	91	38	16	119	101	36	21	127	101	36	21
1a Rolling	67	55	23	5	93	76	34	12	107	89	38	16	120	100	36	21	127	100	36	21
2a Recursive	67	55	23	5	93	76	34	12	107	89	38	16	120	100	36	21	127	100	36	21
2a Rolling	67	55	23	5	91	77	33	12	106	90	37	16	119	100	36	20	126	100	36	20
3a Recursive	37	37	14	6	41	42	24	10	44	45	25	10	46	45	22	8	50	45	22	8
3a Rolling	36	37	24	5	53	52	36	9	61	61	37	10	68	65	33	8	71	65	33	8
1b Recursive	67	55	23	5	93	76	34	11	102	92	38	15	118	101	34	17	127	101	34	17
1b Rolling	67	55	23	5	93	76	34	11	102	92	38	15	118	101	34	17	127	101	34	17
2b Recursive	69	53	23	5	92	77	33	11	100	91	37	15	118	100	34	17	127	100	34	17
2b Rolling	69	53	23	5	92	77	33	11	100	91	37	15	118	100	34	17	127	100	34	17
3b Recursive	39	35	21	6	46	39	30	10	54	43	30	10	56	48	29	8	61	48	29	8
3b Rolling	39	39	26	5	57	56	38	9	64	62	42	10	70	65	37	8	75	65	37	8

The table illustrates the number of times, the rule strictly beats the Passive Buy and Hold strategy for the respective forecasting model and denominator. The simulation was attempted 160 times. For the 1 year horizon, the rule beats the Buy and Hold 67 times for the AR(p) recursive model. As the horizon expands, the number of times the model beats the Buy and Hold increases as well. For denominator C, the success rates diminish after 3 years of horizons.

Table A.13:

Returns on Holding Positions.

Table A.13 shows the accumulated returns if the holding strategy involved deciding at time t the position to be held and then holding onto the position for the whole horizon $t + k$. The results show that the rule still has smaller returns than the Buy and Hold

	1a Rec	1a Rol	1b Rec	1b Rol	2a Rec	2a Rol	2b Rec	2b Rol	Buy and Hold
Denominator A									
6 Years	0.381	0.392	0.386	0.397	0.378	0.394	0.395	0.394	0.46
5 Years	0.318	0.332	0.319	0.335	0.318	0.328	0.327	0.328	0.38
4 Years	0.252	0.262	0.251	0.265	0.250	0.255	0.255	0.255	0.31
3 Years	0.174	0.181	0.174	0.182	0.180	0.185	0.185	0.185	0.23
2 Years	0.116	0.118	0.114	0.122	0.117	0.119	0.120	0.119	0.15
Denominator B									
6 Years	0.374	0.365	0.378	0.365	0.376	0.361	0.386	0.361	0.46
5 Years	0.311	0.307	0.314	0.307	0.327	0.310	0.333	0.311	0.38
4 Years	0.239	0.229	0.242	0.229	0.258	0.245	0.265	0.245	0.31
3 Years	0.170	0.156	0.172	0.156	0.189	0.167	0.191	0.166	0.23
2 Years	0.117	0.097	0.116	0.097	0.129	0.112	0.129	0.112	0.15
Denominator C									
6 Years	0.301	0.293	0.301	0.296	0.342	0.345	0.377	0.334	0.46
5 Years	0.265	0.258	0.265	0.265	0.289	0.281	0.320	0.278	0.38
4 Years	0.194	0.190	0.194	0.196	0.216	0.201	0.238	0.205	0.31
3 Years	0.130	0.126	0.130	0.135	0.147	0.140	0.163	0.141	0.23
2 Years	0.095	0.089	0.095	0.095	0.098	0.098	0.108	0.101	0.15
Denominator D									
6 Years	0.304	0.304	0.304	0.304	0.312	0.312	0.312	0.312	0.46
5 Years	0.247	0.247	0.247	0.247	0.251	0.251	0.251	0.251	0.38
4 Years	0.180	0.180	0.180	0.180	0.182	0.182	0.182	0.182	0.31
3 Years	0.124	0.124	0.124	0.124	0.126	0.126	0.126	0.126	0.23
2 Years	0.078	0.078	0.078	0.078	0.079	0.079	0.079	0.079	0.15

A.1.9 Rolling and Recursive Tests of Stationarity

The dividend growth and earnings growth models (Fama and French (2002)) assume that if Price and Dividends (Earnings) are cointegrated, the sum of the historical averages of dividend-price ratio and dividend(earnings) growth would yield expected returns measures. This is similar to assessing whether the price-dividend ratio and price-earnings ratio are stationary. These tests form the basis of discount rates C and D.

ADF test

For real time purposes, we tested for the stationarity of $\frac{D}{P}$ and $\frac{E}{P}$ using both rolling and recursive ADF tests. The results tend to differ as to which criterion is used to select the residual term in the ADF equation. The various criteria used are Akaike Information Criteria, Bayesian Information Criteria, Schwartz Information Criteria, Hannan Quinn and the Modified Information criteria. We report the Modified Akaike criterion only.

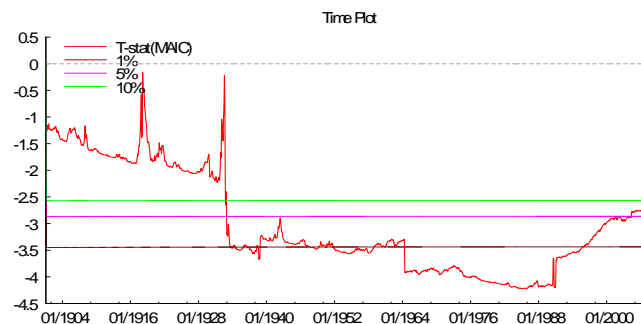


Figure A.9: Recursive ADF T-stat with MAIC Optimal Lag Selection for Price-dividend Ratio.

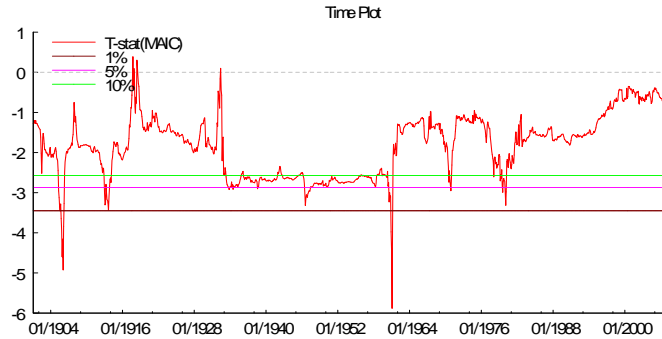


Figure A.10: **Rolling ADF T-stat with MAIC Optimal Lag Selection for Price-dividend Ratio.**

Figures A.9 and A.10 illustrate the recursive and rolling Augmented Dickey Fuller (ADF) test where the number of the lagged error terms is chosen by the Modified Akaike Information criteria. The graph shows the level of rejection subject to the sample under consideration. The horizontal lines show the critical values for rejection at the 1 %, 5 % and 10 % level of significance. In the first 30 years, both the Rolling and Recursive estimate do not tend to reject the null hypothesis of nonstationarity. However, as the sample size is increased for the recursive tests, the null hypothesis is rejected till the end of the sample. In the recent financial crisis the test marginally rejects the null hypothesis at the 5 % level. On the other hand, the rolling window estimation shows that the price-dividend ratio is stationary between 1925 until 1960. A similar pattern to the beginning of the sample is then witnessed from then on. The point this graph conveys is that stationarity may not be seen in a global (whole sample size) but rather in a local context.

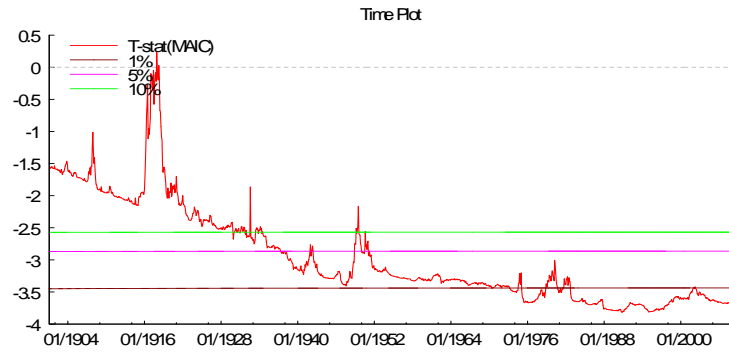


Figure A.11: Recursive ADF T-stat with MAIC Optimal Lag Selection for the Price-earnings Ratio.

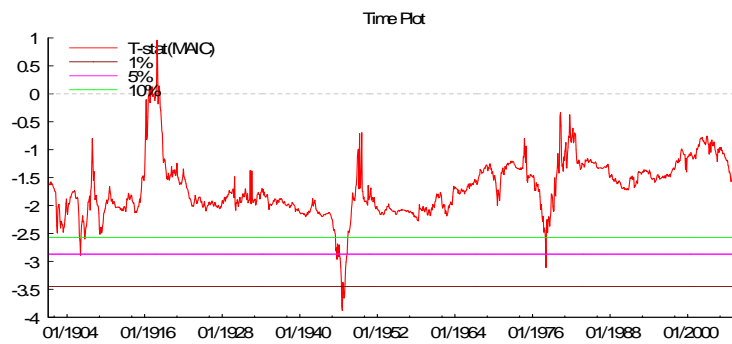


Figure A.12: Rolling ADF T-stat with MAIC Optimal Lag Selection for Price-earnings Ratio.

Figures A.11 and A.12 shows the recursive and rolling ADF test for the price-earnings ratio. Interestingly, the graphical plots are similar at the start of the sample. The rolling window ADF test shows that the price-earnings ratio is mostly nonstationary and is stationary interestingly in two cases, 1944 and 1977. However the recursive window shows that the price-earnings is stationary as the sample size increases.

A.2 Appendix 2

A.2.1 The Present Value Model.

Equations (5.1) , (5.2) and (5.3) are shown again:

$$r_t = \log\left(\frac{P_{t+1} + D_{t+1}}{P_t}\right)$$

$$PD_t = \frac{P_t}{D_t}$$

$$\Delta d_{t+1} = \log\left(\frac{D_{t+1}}{D_t}\right)$$

The return process can be written as:

$$\begin{aligned} r_t &= \log\left(\left(\frac{P_{t+1} + D_{t+1}}{P_t}\right) \cdot \frac{D_t}{D_t} \cdot \frac{D_{t+1}}{D_{t+1}}\right) \\ &= \log\left(\left(\frac{P_{t+1}D_t + D_{t+1}D_t}{P_tD_{t+1}}\right) \cdot \frac{D_{t+1}}{D_t}\right) \\ &= \log\left(\left(\frac{D_t}{P_t} \cdot \frac{P_{t+1}}{D_{t+1}} + \frac{D_t}{P_t}\right) \cdot \frac{D_{t+1}}{D_t}\right) \\ &= \log\left(\left(\frac{P_{t+1}}{D_{t+1}} + 1\right) \frac{D_{t+1}}{D_t} \cdot \frac{D_t}{P_t}\right) \\ &= \log(1 + e^{pd_{t+1}}) + \Delta d_{t+1} - pd_t \end{aligned}$$

Assuming the log linearization of Campbell and Shiller (1988) the returns can be written as

$$r_t \simeq \log((1 + e^{pd_{t+1}})) + \frac{\exp(\overline{pd})}{1 + \exp(\overline{pd})} + \Delta d_{t+1} - pd_t \quad (\text{A.5})$$

$$r_t = \kappa + \rho pd_{t+1} + \Delta d_{t+1} - pd_t \quad (\text{A.6})$$

where $\kappa = \log((1 + e^{pd_{t+1}}))$ and $\rho = \frac{\exp(\overline{pd})}{1 + \exp(\overline{pd})}$

Hence,

$$pd_t = \kappa + \rho pd_{t+1} + \Delta d_{t+1} - r_{t+1} \quad (\text{A.7})$$

(A.7) is the Campbell-Shiller approximation of present value.

A.2.2 State Space Model assuming AR(1)

In this section I reproduce the Kalman Filter used in KVB.

There are two transition equations, one governing the dividend growth rate and the other one governing the mean return:

$$\hat{g}_{t+1} = \gamma_1 \hat{g}_t + \varepsilon_{t+1}^g \quad (\text{A.8})$$

$$\hat{\mu}_{t+1} = \delta_1 \hat{\mu}_t + \varepsilon_{t+1}^\mu \quad (\text{A.9})$$

the two measurement equations are given by :

$$\Delta d_{t+1} = \gamma_0 + \hat{g}_t + \varepsilon_{t+1}^d \quad (\text{A.10})$$

$$pd_t = A - B\hat{\mu}_t + B\hat{g}_t \quad (\text{A.11})$$

(A.11) can be rearranged into (A.9) such that there are only two measurement equations and only one state space model.

$$\hat{g}_{t+1} = \gamma_1 \hat{g}_t + \varepsilon_{t+1}^g \quad (\text{A.12})$$

$$\Delta d_{t+1} = \gamma_0 + \hat{g}_t + \varepsilon_{t+1}^d \quad (\text{A.13})$$

$$pd_{t+1} = (1 - \delta_1)A - B_2(\gamma_1 - \delta_1)\widehat{g}_t + \delta_1pd_t - B_1\varepsilon_{t+1}^\mu + B_2\varepsilon_{t+1}^g \quad (\text{A.14})$$

$$A \text{ is equal to } \frac{\kappa}{1-\rho} + \frac{\gamma_0 - \delta_0}{1-\rho}, B_1 = \frac{1}{1-\rho\delta_1}, B_2 = \frac{1}{1-\rho\gamma_1}.$$

The state equation is defined by:

$$X_{t+1} = FX_t + R\varepsilon_t$$

$$Y_t = M_0 + M_1Y_{t-1} + M_2X_t$$

$$\text{where } Y_t = \begin{bmatrix} \Delta d_t \\ pd_t \end{bmatrix}$$

The variables of the transition equation are X_t and ε_{t+1}^x . They are made up of the following elements:

$$X_t = \begin{bmatrix} \widehat{g}_{t-1} \\ \varepsilon_t^D \\ \varepsilon_t^g \\ \varepsilon_t^\mu \end{bmatrix} \quad F = \begin{bmatrix} \gamma_1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad R = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\varepsilon_{t+1}^x = \begin{bmatrix} \varepsilon_{t+1}^D \\ \varepsilon_{t+1}^g \\ \varepsilon_{t+1}^\mu \end{bmatrix}$$

The parameters of the measurement equation include parameters of the present value model to be estimated. These are defined as:

$$M_0 = \begin{bmatrix} \gamma_0 \\ (1 - \delta_1) * A \end{bmatrix} \quad M_1 = \begin{bmatrix} 0 & 0 \\ 0 & \delta_1 \end{bmatrix}$$

$$M_2 = \begin{bmatrix} 1 & 1 & 0 & 0 \\ B_2(\gamma_1 - \delta_1) & 0 & B_2 & -B_1 \end{bmatrix}$$

The variance covariance matrix from the state space model is given by:

$$\Sigma = \text{var} \begin{bmatrix} \varepsilon_{t+1}^g \\ \varepsilon_{t+1}^\mu \\ \varepsilon_{t+1}^d \end{bmatrix} = \begin{bmatrix} \sigma_g^2 & \sigma_{g\mu} & \sigma_{gd} \\ \sigma_{g\mu} & \sigma_\mu^2 & \sigma_{D\mu} \\ \sigma_{gd} & \sigma_{D\mu} & \sigma_D^2 \end{bmatrix}$$

The Kalman Filter equations are given by the following:

$$\begin{aligned} X_{0|0} &= E[X_0] \\ P_{0|0} &= E[X_0 X_0'] \\ X_{t|t-1} &= F X_{t-1|t-1} \\ P_{t|t-1} &= F P_{t-1|t-1} F' + R \Sigma R' \\ \eta_t &= Y_t - M_0 - M_1 Y_{t-1} - M_2 X_{t|t-1} \\ S_t &= M_2 P_{t|t-1} M_2' \\ K_t &= P_{t|t-1} M_2' S_t^{-1} \\ X_{t|t} &= X_{t|t-1} + K_t \eta_t \\ P_{t|t} &= (I - K_t M_2) P_{t|t-1} \end{aligned}$$

The likelihood function is given by :

$$L = - \sum_{t=1}^T \log(\det(S_t)) - \sum_{t=1}^T \eta_t' S_t^{-1} \eta_t$$

A.2.3 Other Statistical results

Table A.14 shows that differences in the stationarity benchmarks for realized returns and expected returns. Realized returns seem to be a stationary series while expected returns appear to be nonstationary. The test of I(1) tends to be weak for the case of expected returns. It tends not to reject the null. The p-values for the Augmented Dickey Fuller test, Phillips Perron and Dickey Fuller generalized least squares tend to show mixed results. In the case of the Phillips Perron test, the computed statistic is rejected at the 10 % level. In

Table A.14:

Tests of Stationarity.

The table reports stationarity and nonstationarity tests. The first three tests have for null hypothesis that the series is integrated of order zero. The figures in brackets are the p-values and should be read as rejection below the bounds.

	R_t	μ_t	Δd_t	g_t
Stationarity test of I(0)				
Robinson-Lobato	9.54(0)	5.327(0)	1.263(0.103)	-0.676(0.751)
KPSS test	0.574(< 0.02)	0.878(< 0.01)	0.054(< 1)	0.049(< 1)
Lo's RS	1.316(< 0.4)	1.386(0.3)	0.971(< 0.9)	0.962(< 0.9)
Stationarity test of I(1)				
Augmented Dickey-Fuller	-25.34(< 0.01)	-2.199(< 0.9)	-17.07(< 0.01)	-11.85(< 0.01)
Phillips-Perron	-25.34(< 0.01)	-2.601(< 0.1)	-16.99(< 0.01)	-11.75(< 0.01)
DF-GLS	-25.34(< 0.01)	-2.201(< 1)	-16.74(< 0.01)	-11.76(< 0.01)

the case of the other two statistics, the test is rejected at the 90% level. Most of the long memory tests tend to point to the direction that expected returns may be fractionally integrated.

Table A.15:

Correlation of Realized and Expected Dividend Growth.

The table illustrates the correlation of realized and expected dividend growth rates at specific lags.

Lag	Δd_t vs. Δd_{t-j}	g_t vs. g_{t-j}	Δd_t vs. g_{t-j}	g_t vs. Δd_{t-j}
0.	1	1	0.649	0.649
1.	0.636	0.807	0.519	0.962
2.	0.506	0.643	0.420	0.781
3.	0.386	0.520	0.402	0.626
4.	0.387	0.470	0.349	0.489
5.	0.332	0.416	0.319	0.448
6.	0.317	0.374	0.261	0.395
7.	0.258	0.317	0.215	0.366
8.	0.224	0.264	0.160	0.314
9.	0.165	0.208	0.125	0.268
10.	0.128	0.165	0.112	0.212
11.	0.117	0.141	0.092	0.167
12.	0.103	0.116	0.089	0.144

Realized and expected dividends are positively correlated over time. This explains some of the findings outlined in chapter five where it was found that the filtered expected dividend growth predicts realized growth. As the horizon grows, the level of correlation decreases. It is interesting to see that lags of one and two for realized dividend growth rate have higher correlation levels with expected growth.

Table A.16:

Correlation of Realized and Expected Returns.

The table illustrates the correlation of realized and expected returns at different lags as specified in the first column.

Lag	R_t vs R_{t-j}	μ_t vs μ_{t-j}	R_t vs μ_{t-j}	μ_t vs R_{t-j}
0.	1	1	0.255	0.255
1.	0.340	0.992	0.353	0.220
2.	0.097	0.981	0.372	0.211
3.	0.038	0.969	0.365	0.215
4.	0.102	0.960	0.353	0.211
5.	0.155	0.951	0.351	0.205
6.	0.145	0.942	0.356	0.198
7.	0.118	0.931	0.362	0.193
8.	0.113	0.921	0.361	0.189
9.	0.103	0.911	0.357	0.186
10.	0.115	0.901	0.357	0.179
11.	0.097	0.890	0.360	0.176
12.	0.071	0.88	0.359	0.177

Compared to the realized dividend growth, realized returns do not have a high correlation. On the other hand, there is a high level of serial correlation between expected returns and its lags. Even after accounting for 12 months lags, the expected returns at time t is correlated with its twelfth lag at 0.88. There appears to be weak correlation between the returns and lags of expected returns at around approximately 0.35. Moreover, in the case of the expected returns with lags of realized returns, the correlation tends to fall with time from 0.255 till 0.177 for the twelfth lag.

A.2.4 Plot of Trading Rule Returns and Probability Distribution of Relative Prices

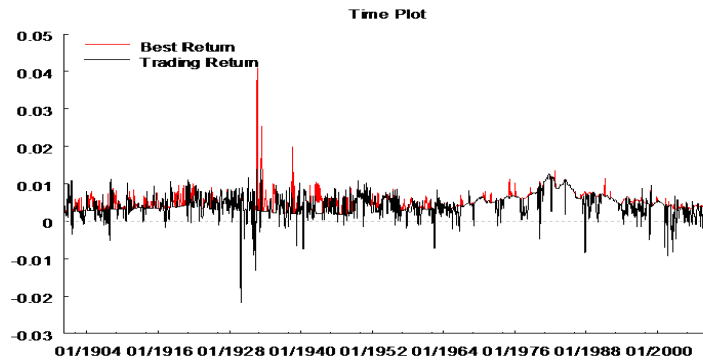


Figure A.13: Plot of Trading Return against Best Return.

Figure A.13 is the plot of the trading return against the maximum return between the risk free asset and equity. The strategy tends to fare worse at the beginning of the sample and at the end of the sample. However, there are no distinct periods when the trading return tends to fare better or worse than the actual returns.

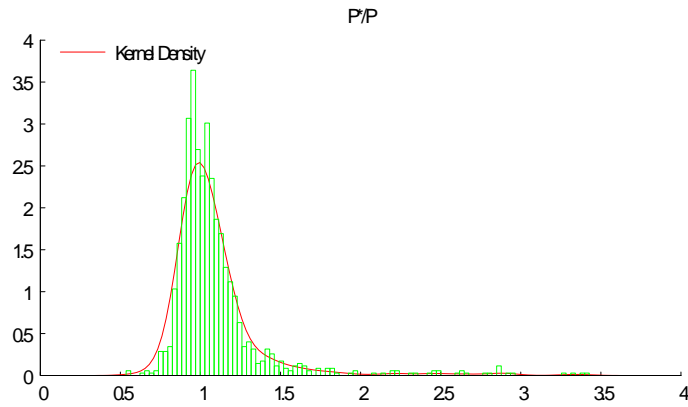


Figure A.14: **Probability Distribution of P_t^*/P_t**

Figure A.14 shows the probability distribution of the present value standardized with price. If the distribution is distributed as a Gaussian distribution with a mean of one, then on average we should expect that the market is under or overpriced. However, the distribution is right skewed implying that the market is overpriced as per the present value.

A.3 Appendix 3

A.3.1 Monte Carlo Experiment on Truncation lags

Table A.17 illustrates the Monte Carlo experiment on the representation of a pure fractional noise process with an $AR(m)$.

Table A.17:

**Monte Carlo experiment on Maximum Likelihood estimates of ‘d’
based on Truncated Autoregressive Models.**

The experiment is performed for $d = 0.05, 0.2$ and 0.4 . The number of Monte Carlo replications were constrained to 100 replications due to the intensive computational cost. $m = 5, 10, 20$ and 40 for a sample sizes of 100 observations and 200 observations.

m	$d = 0.05$	$d = 0.2$	$d = 0.4$
$T = 100$			
AR(5)	0.076	0.212	0.424
AR(10)	0.071	0.207	0.419
AR(20)	0.063	0.206	0.416
AR(40)	0.054	0.199	0.411
$T = 200$			
AR(5)	0.073	0.213	0.437
AR(10)	0.067	0.206	0.414
AR(20)	0.055	0.202	0.403
AR(40)	0.051	0.202	0.402

The Monte Carlo shows that the optimized value of d through the exact likelihood model converges to the true parameter for large m . Large m shows that there is convergence to the true parameter. However, interestingly, we find that for the smaller sample ($T = 100$), the low truncated lags ($m = 5$) have a better approximation than for the longer sample for higher values of d . However, this finding has marginal impact on our own state space model given the truncation level.

A.3.2 Summary Statistics

Table A.18:

**Summary Statistics for Expected and Realized Returns and
Expected and Realized Dividend Growth Rates: 1900-2008**

	\mathbf{R}_t	$\boldsymbol{\mu}_t$	$\boldsymbol{\Delta d}_t$	\mathbf{g}_t
Mean	0.057	0.060	0.015	0.013
Std. Deviation	0.185	0.038	0.114	0.016
Skewness	-0.612	-1.10	-0.710	-0.858
Kurtosis	3.341	3.92	7.805	4.202
Jarque-Bera	7.330	26.07	114.1	19.97

Table A.18 shows the distributional properties of the returns, expected returns, dividend growth and expected dividend growth. Both the mean of the realized and the latent counterpart tend to be very close to each other. The variance of the realized values tends to be higher than that of the filtered values. The skew is larger for the expected values however. It shows that bad news has a marginally higher impact on expectations than realized values.

Table A.19:

Tests of Stationarity on Expected Returns and Expected Dividend Growth Sample 1900-2008.

The Robinson-Lobato, KPSS and Range-scale test assume that the series is stationary under the null hypothesis. The rest of the statistics assume nonstationarity. The figures in brackets show the p-values at which the null hypothesis is rejected.

	\mathbf{R}_t	μ_t	$\Delta \mathbf{d}_t$	\mathbf{g}_t
	Stationarity test of I(0)			
Robinson-Lobato	-0.92{0.82}	0.21{0.42}	-1.28{0.90}	-0.01{0.50}
KPSS test	0.04{< 1}	0.29{< 1}	0.03{< 1}	0.06{< 1}
Lo's RS	0.87{< 0.95}	0.82{< 0.975}	0.77{< 0.99}	0.51{< 0.8}
	Stationarity test of I(1)			
Augmented Dickey-Fuller	-9.20{< 0.01}	-1.77{< 0.9}	-8.89{< 0.01}	-8.64{< 0.01}
Phillips-Perron	-9.30{< 0.01}	-1.79{< 0.9}	-8.92{< 0.01}	-8.72{< 0.01}
DF-GLS	-7.58{< 0.01}	-1.80{< 1}	-8.78{< 0.01}	-8.62{< 0.01}

Similar to the results from table A.14, expected returns is non-stationary. There is stronger rejection from the Phillips-Perron test. The dividend growth and expected dividend growth tends to be stationary, as shown by the Augmented Dickey-Fuller, Phillips-Perron and Dickey-Fuller Generalized Least Squares. However, the stationarity tests, such as the Robinson-Lobato test, tend to show that there is a possibility of returns and expected returns also being $I(0)$.

Table A.20:

Correlation between Expected and Realized Returns up to 5 lags.

Lag	R_t vs. R_{t-j}	μ_t vs. μ_{t-j}	R_t vs. μ_{t-j}	μ_t vs. R_{t-j}
0	1	1	0.197	0.197
1	0.060	0.9208	0.199	-0.150
2	-0.186	0.861	0.142	-0.105
3	0.080	0.823	0.173	-0.076
4	-0.058	0.768	0.185	-0.139
5	-0.108	0.711	0.146	-0.119

Table A.20 shows the correlation over time for returns and expected returns. Realized returns and lags of expected returns tend to be low and the correlation tends to decrease over time from 0.197 to 0.146 in the fifth lag. On the other hand, expected returns and lags of returns tend to be negatively correlated. The autocorrelation of expected returns tends to be high and appears to diminish at a slow rate.

Table A.21:

Correlation between Expected and Realized Dividend Growth up to 5 lags.

Lag	Δd_t vs Δd_{t-j}	g_t vs g_{t-j}	Δd_t vs g_{t-j}	g_t vs Δd_{t-j}
0	1	1	0.186	0.186
1	0.155	0.175	-0.094	0.616
2	0.229	-0.125	-0.047	0.165
3	0.0797	0.075	-0.051	-0.119
4	-0.066	-0.047	-0.028	0.020
5	-0.015	-0.070	0.018	-0.092

In the case of dividend growth (table A.21), the correlation of realized growth with lagged expected growth rate tends to be lower. The first autocorrelation and that at lag five tends to be positive however.

A.3.3 Expected Returns

The statistical properties of expected returns are reported in this section. It presents a first pass misspecification test as to the adequacy of the ARFIMA model. The following tables report both the statistical moments of each series under both models as well as some stationarity and nonstationarity tests.

From table A.22, the mean of the expected returns series is lower than that of the realized values. The series is also marginally less volatile but shares the same level of kurtosis. The test of I(1) in expected returns shows conflicting conclusions as to whether the expected returns series is a stationary series. The AR(1) for both samples tends to show that expected returns is a non-stationary series. However the ARFIMA tends to remove the persistence due to long range components, hence making the series closer to an I(0) process.

Table A.22:

**Descriptive Statistics and Stationarity tests on the Expected
Returns series for Dividend data.**

The stationarity tests differ from each other based on the null hypothesis. Tests of $I(0)$ assume that the null hypothesis is that the series is a stationary series. In the case of the Robinson-Lobato(1998) test, the alternative is a fractional process. Tests of $I(1)$ are tests with null hypothesis being an integrated series of $I(1)$. For the $I(0)$ and $I(1)$ tests, the reported p-values are the rejection regions where the test statistic lies.

Price-dividend Ratio/Dividend Growth						
	1926-2008			1946-2008		
	AR(1)	ARFIMA(1,d,0)	r_t	AR(1)	ARFIMA(1,d,0)	r_t
Mean	0.062	0.064	0.080	0.063	0.065	0.073
Std. Dev	0.049	0.054	0.195	0.045	0.036	0.169
Kurtosis	3.310	3.49	2.944	3.309	4.282	3.090
Test of I(0)						
Robinson-Lobato	-0.075	1.152	-0.599	-	-0.484	0.208
P-value	0.524	0.125	0.726	-	0.686	0.42
KPSS	1.378	0.583	0.073	0.251	0.037	0.080
P-value	0.3	0.025	1	1	1	1
Test of I(1)						
ADF test	-1.11	-4.63	-8.49	-1.214	-2.672	-6.733
P-value	0.9	0.01	0.01	0.9	0.1	0.01
Phillips-Perron	-1.61	-5.81	-8.58	-1.331	-2.800	-6.823
P-value	0.9	0.01	0.01	0.9	0.1	0.01

Table A.23:

Descriptive Statistics and Stationarity tests on the Expected Returns series using Earnings data.

The resulting statistical features of the expected returns are from the optimization model with the price-earnings and earnings growth.

Price-earnings Ratio/Earnings growth						
	1926-2008			1946-2008		
	AR(1)	ARFIMA(1,d,0)	r_t	AR(1)	ARFIMA(1,d,0)	r_t
Mean	0.087	0.081	0.080	0.096	0.060	0.073
Std. Dev	0.079	0.036	0.195	0.064	0.034	0.169
Kurtosis	4.691	3.608	2.944	4.297	4.069	3.090
Test of I(0)						
Robinson-Lobato Test	-0.120	0.537	-0.599	-0.299	-0.139	0.208
P-value	0.548	0.295	0.726	0.618	0.592	0.42
KPSS test	0.457	0.129	0.073	0.576	0.549	0.080
P-value	0.1	1	1	0.025	0.05	1
Test of I(1)						
ADF test	-1.009	-4.46	-8.49	-0.953	-1.700	-6.733
P-value	0.9	0.01	0.01	0.9	0.9	0.01
Phillips-Perron	-2.66	-4.53	-8.58	-2.207	-2.053	-6.823
P-value	0.1	0.01	0.01	0.9	0.9	0.01

The findings from the earnings growth model and the dividend growth model are similar. In both cases, the ARFIMA removes the long range dependence components. The only exception is the 1946-2008 sample. In the latter, expected returns are shown to be nonstationary series. As referred in text, this may be due to the fact that the model is optimized using a small sample size. It is also worth noticing that the standard p-values for the $I(0)$ tests show that the series is not $I(0)$ in the case of the ARFIMA models, which implicitly implies that expected returns are fractionally integrated. Another important statistic worthy of notice is that the kurtosis for the expected series tends to be higher than usual.

A.3.4 Expected Dividend (Earnings) Growth

The distributional properties and stationarity tests of expected dividend (earnings growth properties are presented in tables A.24 and A.25.

Table A.24:

Descriptive Statistics and Stationarity tests on Expected Dividend Growth rate.

	Price-dividend ratio/ Dividend Growth					
	1926-2008			1946-2008		
	AR(1)	ARFIMA(1,d,0)	Δd_t	AR(1)	ARFIMA(1,d,0)	Δd_t
Mean	0.021	0.018	0.017	0.020	0.032	0.021
Std. Dev	0.017	0.045	0.106	0.029	0.023	0.055
Kurtosis	3.175	5.49	10.06	5.773	2.687	5.132
Test of I(0)						
Robinson-Lobato Test	-0.498	-0.294	0.793	0.246	0.723	0.719
P-value	0.691	0.616	0.786	0.403	0.235	0.236
KPSS test	0.322	0.069	0.034	0.066	0.147	0.113
P-value	1	1	1	1	1	1
Test of I(1)						
ADF test	-5.958	-6.582	-7.471	-5.158	-3.289	-4.811
P-value	0.01	0.01	0.01	0.01	0.025	0.01
Phillips-Perron	-8.167	-7.42	-7.474	-5.243	-3.344	4.887
P-value	0.01	0.01	0.01	0.01	0.025	0.01

The ARFIMA(1,d,0) model for the sample 1946-2008 again illustrates some conflicting findings. First, it illustrates a mean of 1 % higher than the other series. The kurtosis is lower at 2.7. However in terms of stationarity tests, the estimated models are parsimonious. The null hypothesis of expected dividend growth rate being I(0) is not rejected at all levels. The same can be said of the I(1) tests, where all the models tend to reject the null hypothesis of nonstationarity.

Table A.25:

Descriptive Statistics and Stationarity tests on Expected Earnings Growth rate.

	Price-earnings Ratio/ Earnings growth					
	1926-2008			1946-2008		
	AR(1)	ARFIMA(1,d,0)	Δe_t	AR(1)	ARFIMA(1,d,0)	Δe_t
Mean	0.087	0.081	0.007	0.012	0.024	0.013
Std.Dev	0.079	0.036	0.226	0.069	0.012	0.217
Kurtosis	4.69	3.574	6.613	6.238	4.451	8.899
Test of I(0)						
Robinson-Lobato Test	-0.119	0.515	-1.395	-0.498	0.165	-1.536
P-value	0.548	0.303	0.918	0.691	0.434	0.938
KPSS test	0.159	0.013	0.049	0.122	0.093	0.152
P-value	1	1	1	1	1	1
Test of I(1)						
ADF test	-2.213	-4.463	-6.486	-4.739	-2.761	-4.698
P-value	0.9	0.01	0.01	0.01	0.1	0.01
Phillips-Perron	-2.66	-4.532	-6.506	-7.327	-3.708	-5.034
P-value	0.05	0.01	0.01	0.01	0.01	0.01

In the case of earnings growth, the results are different across sample sizes. The 1926-2008 sample shows that the expected dividend growth is higher than that of the realized value. One reason which could account for this is the variance of the earnings growth. The observed earnings growth is found to be very volatile, which makes the search for optimal parameters in the state space more daunting. On the other hand, the volatility of the expected earnings growth series tends to be lower. With regards to stationarity test, dividend growth is found to be stationary most of the time.

A.3.5 Correlation over time

The following tables report the correlation coefficient of the different series for both the autoregressive and fractionally autoregressive series over the different samples.

Table A.26:

Correlation of Expected Returns from AR(1) specification.

The table shows the correlation coefficient from the expected returns from the dividend (pd) and earnings (pe) specifications. The first period (*I*) denotes the sample period 1926-2008, while the second period(*II*) involves the period 1946-2008.

	$\mu_t^{PD,I}$	$\mu_t^{PE,I}$	$\mu_t^{PD,II}$	$\mu_t^{PE,II}$
$\mu_t^{PD,I}$	1	0.85	0.99	0.84
$\mu_t^{PE,I}$	0.85	1	0.85	0.99
$\mu_t^{PD,II}$	0.99	0.85	1	0.85
$\mu_t^{PE,II}$	0.84	0.99	0.85	1

The correlation of the expected returns is a naive test as to whether the model is stable. The correlation among the different series is high and hence they reflect that the expected returns from the different state space models follow the same path. The AR(1) process tends to show that there is high correlation over time. For instance, there is a 0.99 correlation between the sample 1926-2008 and 1946-2008. However, this correlation tends to fall when the dividend growth based expected returns series is compared to that of the earnings growth on two different samples. The correlation is high in both cases which adds to the reliability of the AR(1) model for modeling expected returns.

Compared to the AR(1) model, the ARFIMA(1,d,0) performs poorly. However the correlation coefficient is still positive. The highest correlation is 0.88 which involves the correlation between the price-dividend ratio and the price-earnings ratio for the sample 1946-

Table A.27:

Correlation of Expected Returns from ARFIMA(1,d,0) specifications.

	$\mu_t^{PD,I}$	$\mu_t^{PE,I}$	$\mu_t^{PD,II}$	$\mu_t^{PE,II}$
$\mu_t^{PD,I}$	1	0.88	0.56	0.34
$\mu_t^{PE,I}$	0.88	1	0.26	0.04
$\mu_t^{PD,II}$	0.56	0.26	1	0.85
$\mu_t^{PE,II}$	0.34	0.04	0.85	1

2008. However, when the price-earnings ratio is used to compare samples 1946-2008 with 1926-2008, the correlation is very close to zero. Hence, in this framework, we find that the results are not robust.

A.3.6 Monte Carlo Experiment

The Monte Carlo results from estimating an ARFIMA (1, d , 0) when a Markov Regime is the correct data generating process is presented in table (A.28).

Table A.28 illustrates the distributional properties when a model is assumed to follow a Markov Switching model but is estimated using an ARFIMA. According to theory, expected returns is low during booms and high during slumps. I simulate a Markov switching model where the probability of being in a boom period is 0.8 and the probability of being in a slump is 0.2. These probabilities are chosen so as to reflect current state of the sample. Using 1000 simulations, estimated ARFIMA parameters are highly unstable especially in the case of the mean. Based on the Monte Carlo experiment, ARFIMA models may not be stable when it is applied to small sample sizes with regimes.

Table A.28:

Monte Carlo Experiment for 2 regimes with 83 observations.

In the following experiment, a two regime switching model is set up, with regime 1 having an AR(1) process with 0.06 as intercept term and an autoregressive coefficient of 0.9. The second regime has an autoregressive coefficient of 0.8 with an intercept term of 0.09.

	Mean	SD	Skewness	Kurtosis
Intercept	0.06	0.41	-0.36	8.39
AR1	0.84	0.10	-3.48	25.7
Intercept	-1.46	6.78	0.20	17.9
ARFIMA d	0.47	0.49	-0.08	1.33
AR1	0.39	0.44	-0.01	1.31

A.3.7 Univariate Models

In the following tables, I run the regression from their initial transition equation specifications with the filtered series. These results show how expected returns and dividend (earnings) growth behave individually without taking into account the present value approximation.

Table A.29:

Univariate Regressions: Expected Returns using Dividend Growth specification.

The specification for the AR(1) model is $(1 - \varphi_1 L)\mu_t^{AR(1)} = \varphi_0 + \eta_t$. For the ARFIMA model, the specification is $(1 - \varphi_1 L)^d \mu_t^{ARFIMA(0,d,0)} = \varphi_0 + \eta_t$. NH stands for the Nyblom-Hansen test which tests for excessive variation in the estimated parameters. Neglected ARCH is a test of conditional heteroscedasticity in the residual term.

	1926-2008				1946-2008			
	AR(1)		ARFIMA(1,d,0)		AR(1)		ARFIMA(1,d,0)	
	Coeff	P-val	Coeff	P-val	Coeff	P-val	Coeff	P-val
Intercept	0.002	0.5	0.201	0	0.003	0.48	0.027	0
AR(1)	0.94	0	-0.221	0	0.943	0	0.937	0
ARFIMA(d)			0.541	0			-0.123	0.05
Goodness of Fit	0.87		0.194		0.87		0.651	
Autocorrelation	4.766	0.44	4.135	0.53	4.29	0.51	4.742	0.45
Neglected ARCH	17.02	0	16.51	0	14.01	0.02	9.156	0.103
NH - Joint Parameters	0.554	1	1.265	0.05	0.717	0.2	1.07	0.1
NH - Conditional variance	0.242	1	0.860	0.01	0.447	0.075	0.78	0.01
NH - Intercept	0.244	0.2	0.038	0.01	0.295	0.2	0.128	1
NH - AR1	0.061	1	0.111	1	0.129	1	0.128	1
NH - d			0.155	1			0.209	1

The individual specification tests show that the AR(1) is a better model than the ARFIMA based on the point estimates, goodness of fit and other diagnostic tests. The estimates of the AR(1) process roughly equals the estimates from the Kalman Filter as well. There is only a marginal change in the autoregressive parameter over the two periods. However the intercept term is roughly equal to zero. This may be heavily contrasted to the ARFIMA models. In the case of the first sample, the unconditional mean is equal to 0.201. When the sample is increased by 20 data points, the unconditional mean falls by almost ten times. Based on the earlier results, it was found that correlation between the ARFIMA series is just 0.56. Hence, both series are different. However, interestingly, there are conflicting results which emanate from the autoregressive and fractional parameter. For the sample 1926-2008, the autoregressive parameter is - 0.221, while the fractional parameter is 0.541, which is marginally nonstationary. In 1946-2008, the series tends to show that the autoregressive coefficient is a near unit root but at the same time it is antipersistent. It may be inferred from the individual specification tests that the ARFIMA does not perform very well.

According to the models using the price-earnings and earnings growth, the ARFIMA models again perform worse. In both cases, an antipersistent 'd' is noticed and in the first sample, it is not statistically significant. The corresponding autoregressive parameters are equally as high. The interesting aspect that all models seem to behaving is that there is no evidence of serial correlation in the expected returns series. In the AR(1) for the sample 1926-2008, there might be some serial correlation. The Nyblom-Hansen statistics also show that most of the parameters in the model appear to be stable over time,

The fit of dividend growth appears marginally better than that of expected returns. With the exception of the ARFIMA 1946-2008,

Table A.30:

Univariate Regressions: Expected returns using Earnings Growth specification

	1926-2008				1946-2008			
	AR(1)		ARFIMA(1,d,0)		AR(1)		ARFIMA(1,d,0)	
	Coeff	P-val	Coeff	P-val	Coeff	P-val	Coeff	P-val
Constant	0.011	0.42	0.081	0	0.012	0.44	0.020	0.07
AR(1)	0.834	0.	0.804	0	0.851	0	1.050	0
ARFIMA(d)			-0.272	0.24			-0.111	0.01
Goodness of Fit	0.61		0.364		0.622		0.851	
Autocorrelation	8.86	0.11	3.22	0.67	14.88	0.25	5.112	0.402
ARCH	38.43	0	4.303	0.507	11.49	0.49	5.098	0.404
NH - Joint Parameters	0.823	0.2	0.925	0.2	1.127	0.05	0.968	0.2
NH - Conditional variance	0.501	0.05	0.593	0.05	0.547	0.05	0.467	0.075
NH - Intercept	0.295	0.2	0.102	1	0.444	0.075	0.147	1
NH - AR1	0.203	1	0.049	1	0.585	0.05	0.142	1
NH- d			0.051	1			0.212	1

the dividend growth series appear to have a low autoregressive parameter. The dividend growth for the ARFIMA in the smaller sample seems to state the expected dividend growth is highly persistent and is close to nonstationarity. The Nyblom-Hansen test seems to show that the statistic tends to fluctuate a lot over time for all the parameters. Interestingly, it is the conditional variance of the series which tends to fluctuate the most.

According to the expected earnings growth, both ARFIMA models tend to exhibit high persistence. Similarly, the goodness of fit is relatively high for the ARFIMA ranging from 0.36 to 0.47. However, the goodness of fit of such models cannot be an adequate measure of how good the model performs. Overall, the model performs worse as there is high instability in the expected returns series.

Table A.31:

Univariate Regressions on Dividend Growth.

	1926-2008				1946-2008			
	AR(1)		ARFIMA(1,d,0)		AR(1)		ARFIMA(1,d,0)	
	Coeff	P-val	Coeff	P-val	Coeff	P-val	Coeff	P-val
Constant	0.018	0	0.015	0.01	0.012	0	0.010	0
AR(1)	0.100	0.46	0.191	0	0.382	0.05	0.695	0
Goodness of Fit	0.01		0.036		0.145		0.484	
Autocorrelation	4.157	0.527	3.852	0.571	3.619	0.61	6.156	0.291
ARCH	6.437	0.266	15.32	0	39.86	0	10.49	0.062
NH- Joint Parameters	0.871	0.1	1.154	0.05	1.063	0.05	1.040	0.05
NH- Conditional variance	0.704	0.025	0.927	0.01	0.927	0.01	0.769	0.01
NH - Intercept	0.287	0.2	0.051	1	0.110	1	0.280	0.2
NH - Ar1	0.116	1	0.067	1	0.074	1	0.331	0.2

Table A.32:

Univariate Regressions on Earnings Growth

	Monthly				Annually			
	AR(1)		ARFIMA(1,d,0)		AR(1)		ARFIMA(1,d,0)	
	Coeff	P-val	Coeff	P-val	Coeff	P-val	Coeff	P-val
Constant	0.021	0	0.018	0.12	0.011	0.21	0.025	0
AR(1)	-0.024	0.86	0.601	0	0.042	0.82	0.683	0
Goodness of Fit	0.001		0.359		0.002		0.467	
Autocorrelation	3.627	0.46	5.380	0.25	4.022	0.55	8.241	0.14
ARCH	3.460	0.48	3.215	0.52	4.78	0.44	21.573	0
NH- Joint Parameters	0.410	1	0.669	1	0.597	1	0.955	0.07
NH- Conditional variance	0.266	0.2	0.289	0.2	0.401	0.07	0.843	
NH - Intercept	0.060	1	0.267	0.2	0.094	1	0.071	1
NH - Ar1	0.058	1	0.293	0.2	0.015	1	0.068	1

A.3.8 κ_t from 2 stage Model

The following plots show the time series plot of κ , when it is computed from dividend and earnings.

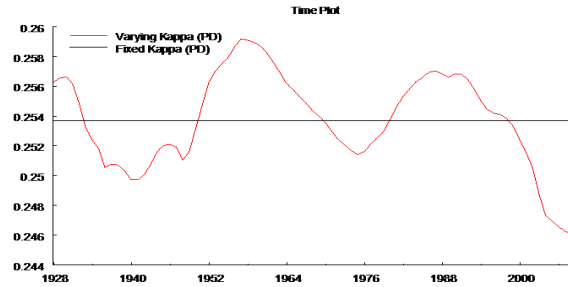


Figure A.15: **Plot of κ_t for Dividend Growth from 1926-2008.** κ_t is defined as $\log(1 + \exp(\overline{pd}_t)) - \rho_t \overline{pd}_t$ where $\rho_t = \frac{\exp(\overline{pd}_t)}{1 + \exp(\overline{pd}_t)}$. κ (non time varying) implies that that this is computed using the whole sample size.

According to the graphical plot A.15, κ_t tends to emulate the behaviour of a business cycle.

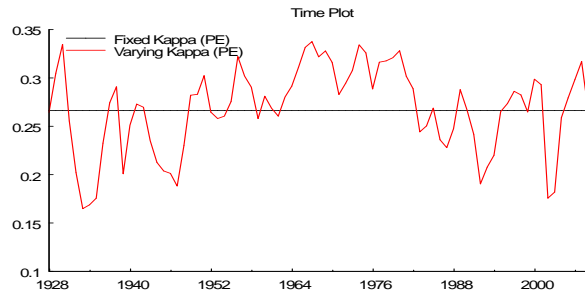


Figure A.16: **Plot of Kappa for Earnings growth from 1926-2008.** κ_t is defined as $\log(1 + \exp(\overline{pe}_t)) - \rho_t \overline{pe}_t$ where $\rho_t = \frac{\exp(\overline{pe}_t)}{1 + \exp(\overline{pe}_t)}$.

Compared to the computed κ_t from the price-dividend ratio, that from earnings (figure A.16) is more volatile. The volatility of the long run equilibrium is very intuitive. It means that in every period, there are mean deviations.

A.3.9 Expected Return

The following graphical plots show the time series plots of the expected returns using the AR(1) and ARFIMA (1,d,0) plots for the dividend and earnings data for the sample 1926-2008 and 1946-2008.

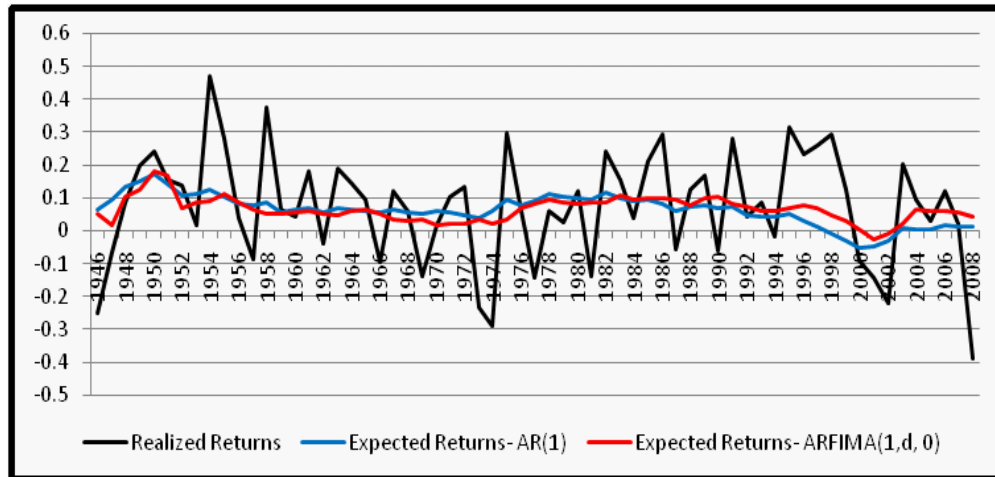


Figure A.17: Plot of Expected Returns for AR(1) and ARFIMA(1,d,0): Sample 1946-2008 using Dividend Data.

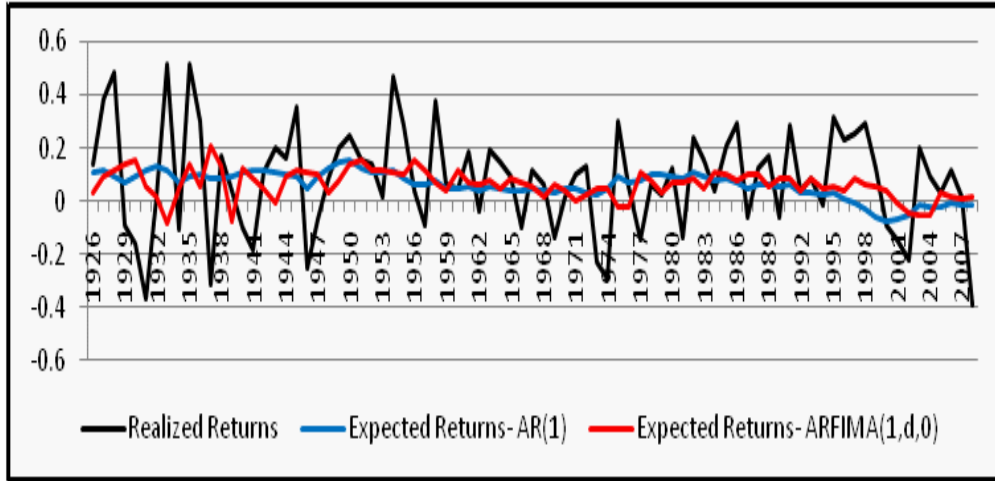


Figure A.18: Plot of Expected Returns for AR(1) and ARFIMA(1,d,0) and Realized Returns for 1926-2008 using Dividend Data.

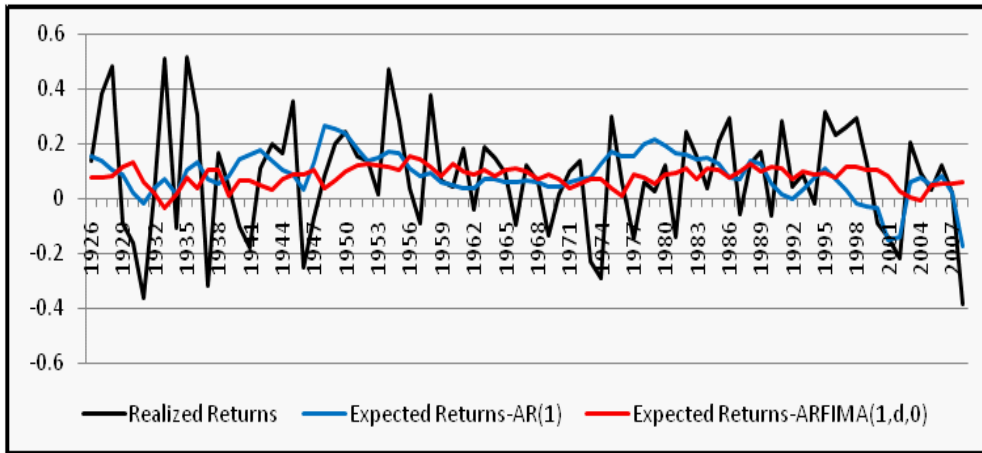


Figure A.19: Plot of Expected Returns for AR(1) and ARFIMA(1,d,0): Sample 1926-2008 using Earnings Data.

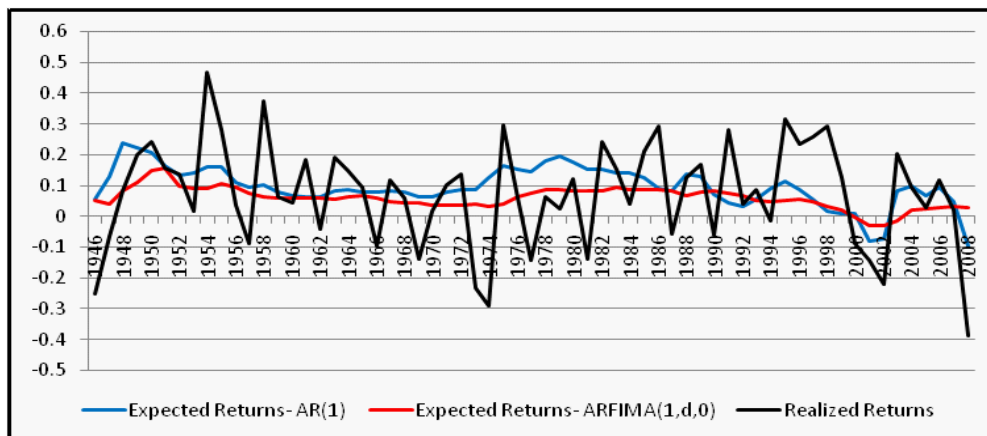


Figure A.20: Plot of Expected Returns for AR(1) and ARFIMA(1,d,0): Sample 1946-2008 using Earnings Data.

The graphical plots show that expected returns tend to be stable over time, irrespective of the specification. However, it is interesting to note that expected returns in the 1946-2008 series tend to be smoother, where as the sample 1926-2008 tends to be more volatile over time. It is also worth noting that the present value approach from the present value does not put bounds on the expected returns being always positive. In all the four graphical plots, the expected returns is negative during the period 1999-2001.

A.3.10 Dividend and Earnings growth

The following graphical plots show the real time estimates of the dividend and earnings growth over time.

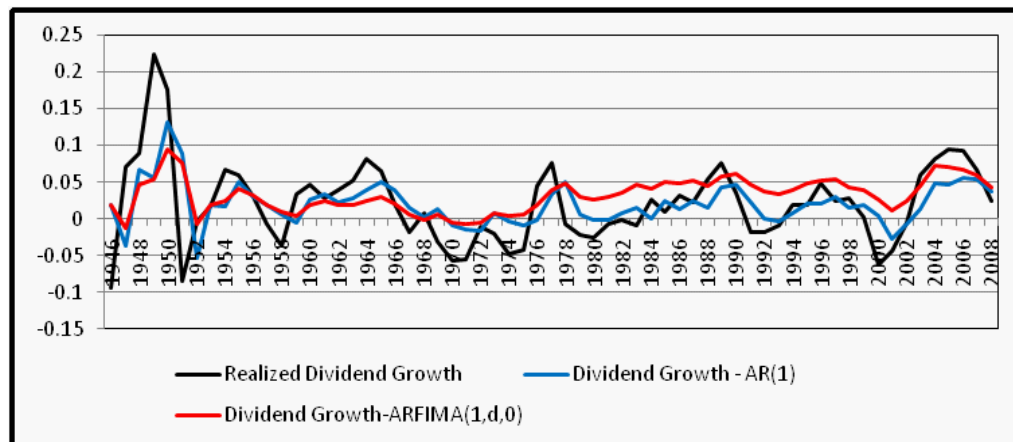


Figure A.21: Plot of Dividend Growth for AR(1) and ARFIMA(1,d,0): Sample 1946-2008.

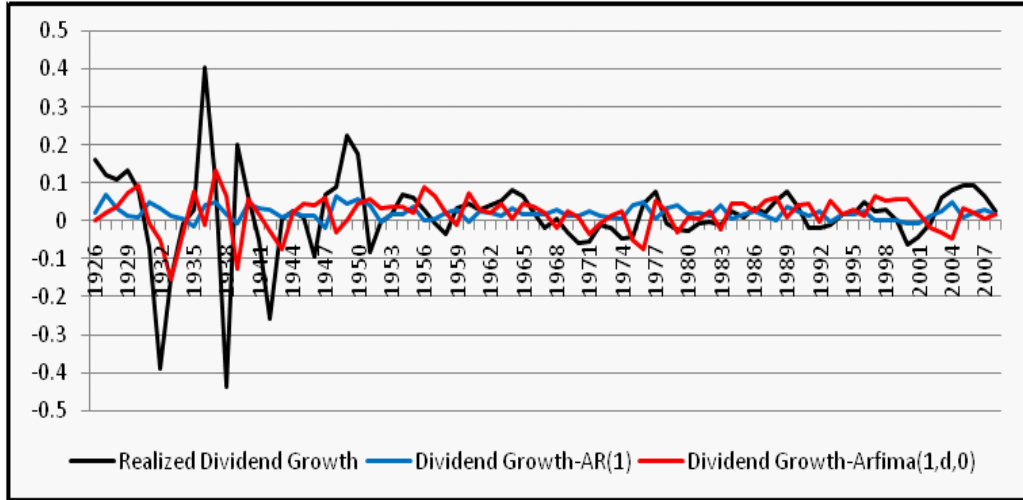


Figure A.22: Plot of Dividend Growth for AR(1) and ARFIMA(1,d,0): Sample 1926-2008.

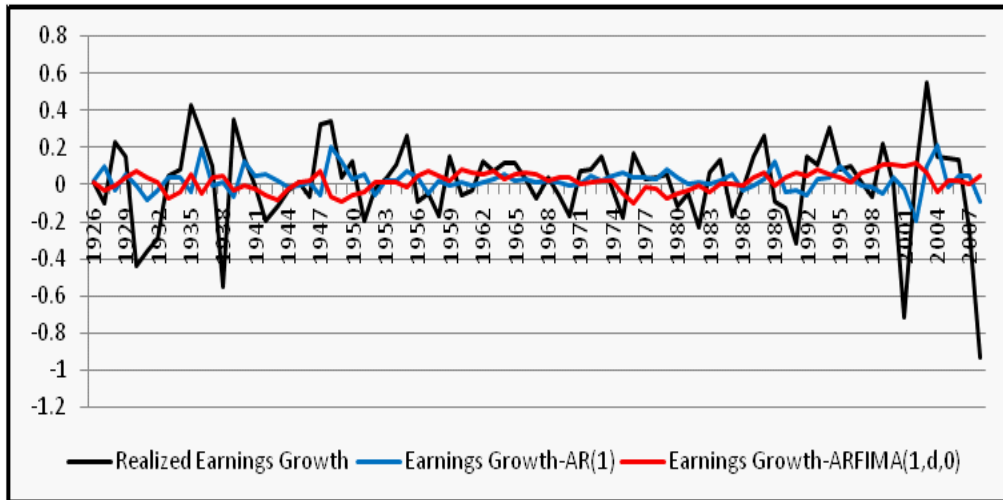


Figure A.23: Plot of Earnings Growth for AR(1) and ARFIMA(1,d,0): Sample 1926-2008.

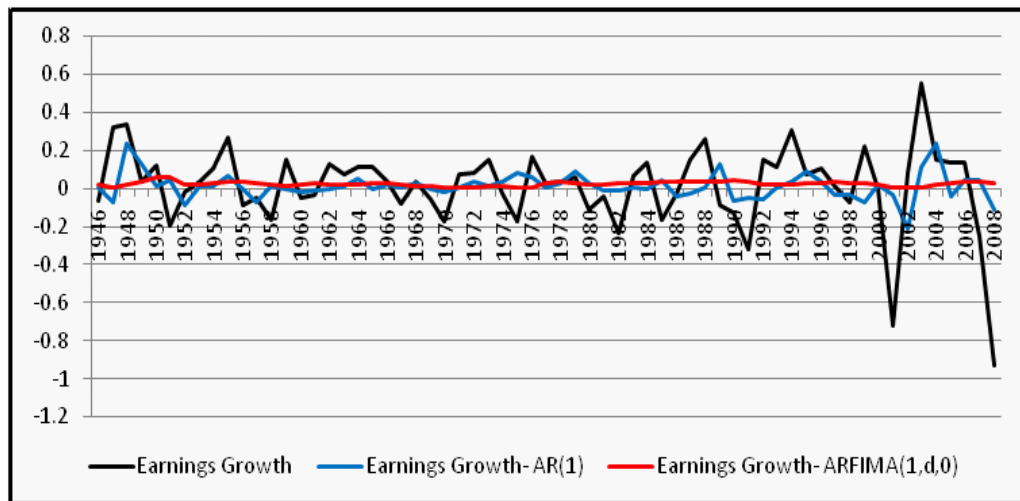


Figure A.24: Plot of Earnings Growth for AR(1) and ARFIMA(1,d,0): Sample 1946-2008.

The graphical plots show varying levels of volatility. The sample 1946-2008 for expected dividend growth (A.21) and earnings growth (A.24) tend to illustrate that the expectations of earnings and dividend from both models tend to be smooth. This is strongly reflected in A.24 for the ARFIMA model. In the 1926- 2008 model, there does not appear to be a lot of variation.

A.3.11 Plot of Trading Returns

Graphical Plots A.25 to A.28 show the evolution of wealth after adopting the trading strategy from the filtered expected returns.

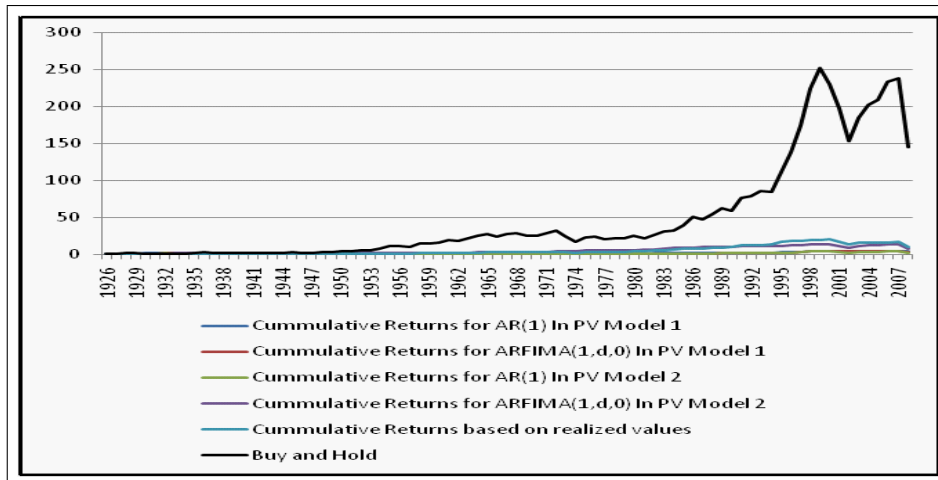


Figure A.25: Cumulated Returns based on the Price-dividend ratio: Sample 1926-2008.

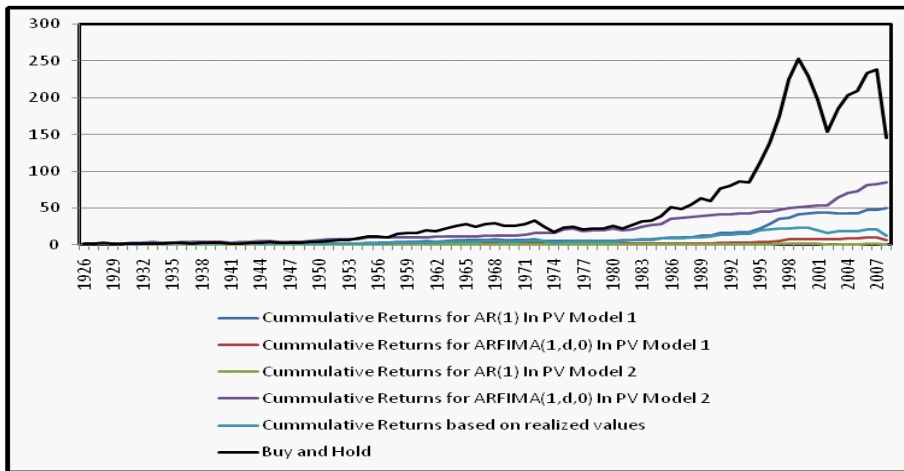


Figure A.26: Cumulated Returns under Price-earnings ratio 1926-2008.

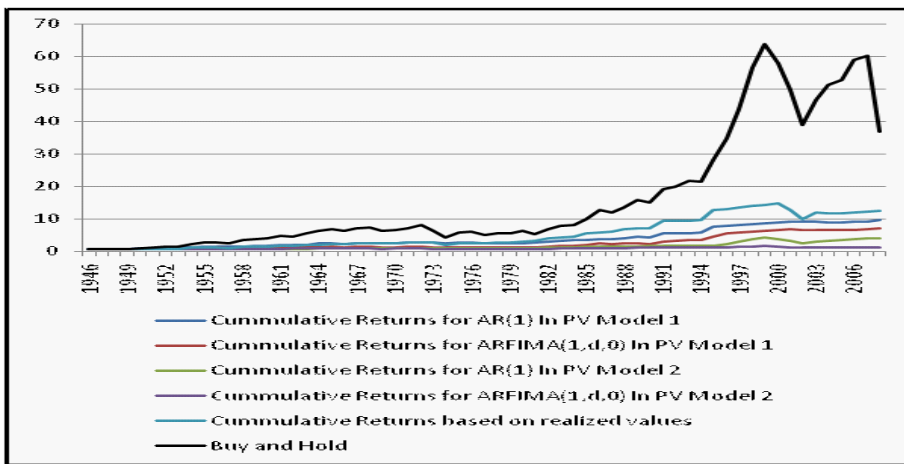


Figure A.27: Cumulated Returns for the Price- dividend ratio: Sample 1946-2008

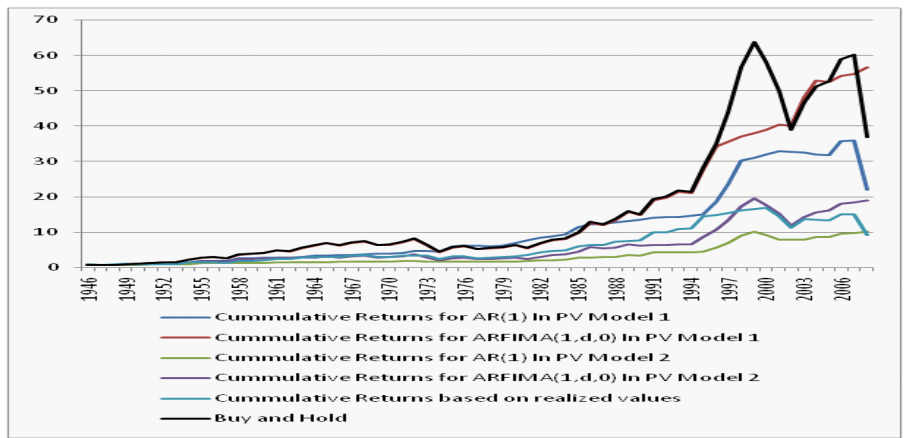


Figure A.28: Cumulated Returns based on the Price-earnings Ratio for the Sample 1926-2008.

It is worthy of mention that nearly all the trading strategies do not appear to beat the Buy and Hold strategy in real time. The only category of expected returns which beats the Buy and Hold is the ARFIMA(1,d,0) for the 1946-2008 earnings growth sample. It is found that by adopting the earnings definition of expected returns, the trading strategy tends to perform better than its counterparts.

A.4 Appendix 4

A.4.1 Proof of Proposition 1

Let γ_k denote the k th autocovariance, defined by the well-known identity

$$\gamma_k = \int_0^{2\pi} \cos(k\omega) f(\omega) d\omega.$$

For the skip-sampled data, the autocovariances are γ_{nk} where

$$\begin{aligned} \gamma_{nk} &= \int_0^{2\pi} \cos(nk\omega) f(\omega) d\omega \\ &= \sum_{j=0}^{n-1} \int_{2\pi j/n}^{2\pi(j+1)/n} \cos(nk\omega) f(\omega) d\omega \\ &= \sum_{j=0}^{n-1} \int_0^{2\pi/n} \cos(nk\omega) f(\omega + 2\pi j/n) d\omega \\ &= \int_0^{2\pi} \cos(k\lambda) f_n(\lambda) d\lambda \end{aligned}$$

where the third equality makes use of the fact that $\cos(nk\omega) = \cos(nk\omega + 2\pi j)$, and the fourth one makes the change of variable $\lambda = n\omega$ and the substitution

$$f_n(\lambda) = \frac{1}{n} \sum_{j=0}^{n-1} f((\lambda + 2\pi j)/n).$$

A.4.2 Proof of Proposition 3

Letting

$$A(\lambda, k, n) = \frac{\cos(\lambda/2n)}{\sin(\lambda/2n)} \sin(\pi k/n) + \cos(\pi k/n)$$

note first that

$$\frac{dH}{dd} = \frac{1}{n} \sum_{k=0}^{n-1} \left[-2A(\lambda, k, n)^{-2d} \frac{g((\lambda + 2\pi k)/n)}{g(\lambda/n)} \log A(\lambda, k, n) + A(\lambda, k, n)^{-2d} \frac{d}{dd} \left(\frac{g((\lambda + 2\pi k)/n)}{g(\lambda/n)} \right) \right].$$

We obtain a formula for the derivative in the second term, and show that this is bounded in the limit. The terms of the form (??) depend on d because the data used to construct the sieve autoregressive estimates are the fractional differences of the measured data. Assume that p is fixed, and let $z_t = (1 - L)^d x_t$ and so let \mathbf{Z}_0 ($T - p \times p$) be the normalized data matrix whose columns are the vectors $\mathbf{z}_j = (z_{p+1-j}, \dots, z_{T-j})'$ for $j = 1, \dots, p$. Also, let \mathbf{Z}_j for $j = 1, \dots, p$ denote the matrix equal to \mathbf{Z}_0 except that the j th column has been replaced by $\mathbf{z}_0 = (z_{p+1}, \dots, z_T)'$. Then, note that the coefficients $\hat{\phi}_j$ in the autoregression of order p can be written using Cramer's rule as

$$\hat{\phi}_j = \frac{|\mathbf{Z}'_0 \mathbf{Z}_j|}{|\mathbf{Z}'_0 \mathbf{Z}_0|}, \quad j = 1, \dots, p.$$

Let these elements define the $p + 1 \times 1$ -vector $\hat{\phi}$ by also putting $\hat{\phi}_0 = -1$.

Now, let $\mathbf{Q}(\theta)$ ($p + 1 \times p + 1$) denote the Fourier matrix with elements $q_{jk} = e^{i\theta(j-k)}$ for $j, k = 0, \dots, p$. Setting $\theta_1 = (\lambda + 2\pi k)/n$ and $\theta_2 = \lambda/n$, note that

$$\frac{\hat{g}(\theta_1)}{\hat{g}(\theta_2)} = \frac{|\hat{\phi}(e^{-i\theta_1})|^{-2}}{|\hat{\phi}(e^{-i\theta_2})|^{-2}} = \frac{\hat{\phi}' \mathbf{Q}(\theta_2) \hat{\phi}}{\hat{\phi}' \mathbf{Q}(\theta_1) \hat{\phi}} = \frac{\mathbf{b}' \mathbf{Q}(\theta_2) \mathbf{b}}{\mathbf{b}' \mathbf{Q}(\theta_1) \mathbf{b}}$$

where \mathbf{b} is the $p + 1$ -vector having elements $b_0 = -T^{-p} |\mathbf{Z}'_0 \mathbf{Z}_0|$ and

$b_j = T^{-p} |\mathbf{Z}'_0 \mathbf{Z}_j|$ for $j = 1, \dots, p$. In this notation we have

$$\frac{d}{dd} \left(\frac{\hat{g}(\theta_1)}{\hat{g}(\theta_2)} \right) = \left[\frac{\mathbf{b}'[\mathbf{Q}(\theta_2) + \mathbf{Q}'(\theta_2)]}{\mathbf{b}'\mathbf{Q}(\theta_1)\mathbf{b}} - \frac{\mathbf{b}'\mathbf{Q}(\theta_2)\mathbf{b}\mathbf{b}'[\mathbf{Q}(\theta_1) + \mathbf{Q}'(\theta_1)]}{[\mathbf{b}'\mathbf{Q}(\theta_1)\mathbf{b}]^2} \right] \frac{d\mathbf{b}}{dd}$$

and it remains to evaluate the second right-hand side factor.

Start with the elements of the \mathbf{Z}_j matrices. Considering row t , let m denote the generic lag associated with a column of \mathbf{Z}_j . Using the argument from Tanaka (1999), Section 3.1, the derivatives with respect to d can be written as

$$\begin{aligned} \frac{dz_{t-m}}{dd} &= \frac{d}{dd} (1-L)^d x_{t-m} \\ &= \log(1-L)(1-L)^d x_{t-m} \\ &= - \sum_{k=1}^{\infty} k^{-1} z_{t-m-k} \\ &= z_{t-m}^* \end{aligned}$$

where the last equality defines z_{t-m}^* . We have from Magnus and Neudecker (1988), p148, that for $j = 0, \dots, p$,

$$\begin{aligned} d |T^{-1} \mathbf{Z}'_0 \mathbf{Z}_j| &= |T^{-1} \mathbf{Z}'_0 \mathbf{Z}_j| \cdot \text{tr}(T^{-1} \mathbf{Z}'_0 \mathbf{Z}_j)^{-1} T^{-1} d(\mathbf{Z}'_0 \mathbf{Z}_j) \\ &= |T^{-1} \mathbf{Z}'_0 \mathbf{Z}_j| \cdot \text{tr}(T^{-1} \mathbf{Z}'_0 \mathbf{Z}_j)^{-1} T^{-1} (d\mathbf{Z}'_0 \mathbf{Z}_j + \mathbf{Z}'_0 d\mathbf{Z}_j) \\ &= |T^{-1} \mathbf{Z}'_0 \mathbf{Z}_j| \cdot \text{tr}(T^{-1} \mathbf{Z}'_0 \mathbf{Z}_j)^{-1} (T^{-1} \mathbf{Z}'_0{}' \mathbf{Z}_j + T^{-1} \mathbf{Z}'_0 \mathbf{Z}_j^*) dd \\ &= b_j^* dd \end{aligned}$$

(defining b_j^*) where the \mathbf{Z}_j^* denote the matrices with elements z_{t-m}^* , with the value of m defined as appropriate, according to the construction of \mathbf{Z}_j . Letting \mathbf{b}^* denote the vector with elements $-b_0^*$ and b_j^* , for $j = 1, \dots, p$, we now have the result

$$\frac{d}{dd} \left(\frac{\hat{g}(\theta_1)}{\hat{g}(\theta_2)} \right) = \frac{\mathbf{b}'[\mathbf{Q}(\theta_2) + \mathbf{Q}'(\theta_2)]\mathbf{b}^*}{\mathbf{b}'\mathbf{Q}(\theta_1)\mathbf{b}} - \frac{\mathbf{b}'\mathbf{Q}(\theta_2)\mathbf{b}\mathbf{b}'[\mathbf{Q}(\theta_1) + \mathbf{Q}'(\theta_1)]\mathbf{b}^*}{[\mathbf{b}'\mathbf{Q}(\theta_1)\mathbf{b}]^2}.$$

Since $\{z_t\}$ is a weakly dependent process by hypothesis, the process

z_t^* is covariance stationary. It follows directly that, for every finite p , \mathbf{b}^* converges in probability to a non-stochastic limit, depending on the autocovariances of $\{z_t\}$. From the fact that \mathbf{b} converges in the same manner, and the Slutsky theorem, the proposition follows under the conditions stated.

Two simplifying assumptions have been made to reach this conclusion. The first is that z_t has been constructed as an infinite order moving average, whereas in practice the sums will be truncated, containing only the first $t - m$ terms. However, since the truncation affects at most a finite number of terms, this cannot change the value of the limit. Also, since z_t is a weakly dependent process by hypothesis, the autocovariances are summable and hence equal zero for lags exceeding some finite value. Letting p tend to infinity with T cannot change the distribution of $B(n, d)$ beyond some point, since the additional elements of \mathbf{b} and \mathbf{b}^* have sums converging to zero as p increases.