

**Artem Abramov**  
**Polina Panfilova**

BI Norwegian Business School – Master Thesis

# Adjusted price-dividend ratio and stock returns predictability

Thesis supervisor: Bruno Gerard

Examination code and name: GRA 1900

Hand-in date:  
02.09.2013

Campus:  
BI Oslo

Programme:  
Master of Science in Financial Economics  
Master of Science in Business and Economics

“This thesis is a part of the MSc programme at BI Norwegian Business School. The school takes no responsibility for the methods used, results found and conclusions drawn.”

---

## Content

<b>1. INTRODUCTION.....</b>	<b>1</b>
<b>2. LITERATURE REVIEW.....</b>	<b>2</b>
<b>3. DATA .....</b>	<b>6</b>
<b>4. PRESENT VALUE MODELS .....</b>	<b>10</b>
4.1 BENCHMARK MODEL .....	11
4.2 ADJUSTED PRICE-DIVIDEND RATIO .....	14
4.2.1 <i>Historical mean adjustment</i> .....	16
4.2.2 <i>ARMA adjustment</i> .....	17
4.2.3 <i>AR adjustment</i> .....	18
4.2.4 <i>How different the adjusted price-dividend ratios are?</i> .....	18
<b>5. EMPIRICAL FRAMEWORK.....</b>	<b>19</b>
5.1 IN-SAMPLE ESTIMATION .....	20
5.1.1 <i>In-sample estimation of the benchmark model</i> .....	20
5.1.2 <i>In-sample estimation of predictive regressions based on adjusted price-dividend ratio</i>	22
5.1.3 <i>Comparison of the models</i> .....	23
5.2 OUT-OF-SAMPLE ESTIMATION .....	24
5.3 ECONOMIC SIGNIFICANCE.....	25
<b>6. ESTIMATION RESULTS .....</b>	<b>27</b>
6.1 IN-SAMPLE ESTIMATION RESULTS .....	27
6.1.1 <i>Benchmark Model</i> .....	28
6.1.2 <i>Predictive regressions</i> .....	29
6.1.3 <i>Comparison of benchmark model with predictive regressions</i> .....	30
6.2 OUT-OF-SAMPLE ESTIMATION RESULTS.....	30
6.3 ECONOMIC SIGNIFICANCE.....	35
<b>7. CONCLUSION .....</b>	<b>39</b>
<b>APPENDIX A .....</b>	<b>41</b>
<b>APPENDIX B .....</b>	<b>43</b>
<b>APPENDIX C .....</b>	<b>44</b>
<b>BIBLIOGRAPHY .....</b>	<b>47</b>

## 1. Introduction

The question of predictability of stock returns has always played an important role in financial economics. Researchers have been concerned with identification of fundamental economic forces, which drive a capital gain process, and implication of various macroeconomic shocks for the equity returns. The question was also very important for ordinary participants of financial markets, because any evidence of predictability would generate feasible trading strategies and lead to better understanding of capital allocation process.

The purpose of this master's thesis is to show how forecasting of future dividends can improve predictability of stock returns for the aggregate stock market. We investigate if estimation of future dividends and dividend growth rates in the spirit of Lacerda and Santa-Clara (2010) leads to better estimates of future equity returns. To do this we replicate their methodology for our dataset focusing on both in- and out-of-sample performance of predictive variables and propose alternative models for dividend growth rate forecasts. The first model is based on the assumptions of Binsbergen and Koijen (2010) and is chosen due to an excellent fit to actual data. Under their assumptions, realized dividend growth rates are ARMA(1,1)-process in case of cash-invested dividends. In other words, we test if our approach leads to the same results as a complex latent-variables model of Koijen and van Binsbergen and if simple OLS technique is sufficient to study stock returns predictability and dividend growth rates jointly. The second predictor is an intermediate case, where we model realized dividend growth rates as the first order autoregressive process.

First, we forecast dividend growth rate and employ it to construct an adjusted price-dividend ratio. Then, we test if this adjustment contributes significantly to predictability of future returns. Moreover, we completely replicate the procedure of Binsbergen and Koijen (2010) to use in-sample performance of their model as a benchmark. We also extend their methodology to evaluate the performance of their model out-of-sample. The model is re-estimated at each time step using only data available by that moment which allows us to construct out-of-sample measure of performance and compare it with our approach.

When on studies stock returns predictability, it is usually difficult to understand practical importance of results at first glance due to typically small gains in predictive power. Therefore, we additionally explicitly investigate if small gains from statistical perspective lead to substantial benefits in capital allocation process by construction of simple trading strategies.

The rest of the master's thesis is organized as follows. In Section 2 we review all relevant literature. Section 3 contains description of the data. In Section 4 we present different versions of log-linearized present-value model and price-dividend ratio adjustments. Section 5 is devoted to empirical transformation of the model and description of estimation techniques. Section 6 demonstrates estimation results and their qualitative implications. In Section 7 we conclude our thesis and discuss potential dimensions of further analysis.

## **2. Literature review**

In 60-70's majority of researchers believed in the efficient market hypothesis implying that predictability of stock returns was considered to be impossible. However, even before a deep study of the question of predictability started, some researchers and practitioners (Dow, 1920 and Ball, 1978) hypothesized that dividend-price ratio (D/P) could be used to forecast stock returns. The intuition behind that hypothesis, which came directly from the famous Dividend Discount Model (DDM) initially proposed by Gordon in 1959, is the following: dividends are high relative to stock prices when expected returns are high, assuming that future dividend growth rates remain constant. Number of researchers, for example, Rozeff (1984), Flood, Hodrick and Kaplan (1986) and Campbell and Shiller (1987) found statistical evidence supporting the hypothesis, but one of the most influential paper was written by Fama in French (1988) in which they not only confirmed statistical significance of dividend yields for prediction of future returns, but also discovered that forecasting power increases with the return horizon and provided strong economical intuition to support their findings. Other fundamental paper was written by Campbell and Shiller (1988). They proposed a log-linearized present-value model, which relates log price-dividend ratio to expectations of log dividend growth rates and log returns at the price of small approximation error. Thus, they stress that, first, variation of price-dividend ratio can reflect not only changes in expected returns, but also changes in expected

dividends. This fact was also emphasized by Menzly, Santos and Veronesi (2004) and Goetzmann and Jorion (1995). Moreover, many researchers, for example, Stambaugh (1999), claim that dividends are highly persistent, implying consequently persistency in dividend yields. This persistency can lead to inaccurate support of the stock returns predictability hypothesis. Thus, one may conclude that stock returns and dividend growth rates are best studied jointly.

Ability to forecast stock returns for the aggregate stock market was also tested for many other valuation ratios. However, what made researchers so focused on dividend yield was simple interpretation of its relationship with stock returns, for example, according to DDM. On the contrary, it is not that easy to find convenient economic arguments, for example, for book-to-market ratio to support the hypothesis that it predicts future returns. This unique feature of price-dividend ratio gave rise to many other papers with different modifications of the present-value model. These were written, for example, by Kojien and van Binsbergen (2010), Lacerda and Santa-Clara (2010), Trojani and Piatti (2012), Cochrane (2007), Lettau and Van Niewerburgh (2008), Pasto and Veronesi (2003), Pastor and Veronesi (2006), Bekaert, Engstrom and Grenadier (2001), Burnside (1998), Ang and Liu (2004), Brennan and Xia (2005) and Rytchkov (2007). In papers written by Pasto and Veronesi (2003), Pastor and Veronesi (2006), Bekaert, Engstrom and Grenadier (2001), Ang and Liu (2004) and Brennan and Xia (2005) price-dividend ratio is presented as an indefinite integral of exponentially-quadratic terms making empirical part of the work much more technical relative to other papers. They employ either generalized method of moments or a two-step iterative procedure to estimate their model. In turn, Kojien and van Binsbergen (2010), Trojani and Piatti (2012), Rytchkov (2007) and Cochrane (2007) combine the present-value model with the assumption that both expected returns and expected dividend growth are latent variables that follow an exogenously-specified time-series model. Then, they assume normality of the shocks to estimate the model with the maximum likelihood and use filtering techniques to uncover expected returns and expected dividend growth rates. Rytchkov (2007) and Cochrane (2007) study methodology construction, derivation of main properties of state-space models, applicability of Kalman filter and relaxation of different assumptions and the consequences for estimation techniques. Kojien and

van Binsbergen (2010) focus on the empirical side of the latent-variables approach and aggregate the whole history of the price-dividend ratio and dividend growth rates to deliver predictors for future returns and dividend growth rates. Since our master's thesis is closely related to the paper of Kojien and van Binsbergen (2010), we proceed to a more detailed discussion of their work.

Kojien and van Binsbergen (2010) model expected returns and expected dividend growth rates as latent variables, which follow low-order autoregressive processes. Following Pastor and Stamabaugh (2006) and Cochrane (2008) they assume that expected returns follow AR(1)-process, however they treat expected dividend growth rates differently depending on the choice of reinvestment strategy. Since they try to avoid effects of seasonality in dividend payments, they consider an annual model, which requires taking into account how dividends received within a particular year are reinvested. Two extreme reinvestment strategies are studied in detail. First, they reinvest dividends in 30-day T-bill and call it cash-invested dividends. Second, they reinvest dividends in the aggregate stock market and refer to them as market-invested dividends. Market-invested dividends appear to be far more volatile than cash-invested dividends supporting the fact that the choice of reinvestment strategy is extremely important. They assume that cash-invested expected growth rates are an AR(1)-process and show that market-invested expected growth rates also exhibit moving average component and follow ARMA(1,1)-process. After specification of processes for latent variables, they employ log-linearization of realized returns in order to connect expected and realized variables through measurement equations. Then, they use Kalman filter not only to estimate unknown parameters, but also to filter out most likely values of latent variables. Later they find what fraction of realized returns and dividend growth variation can be explained by expected values, and compare these numbers to values of regular linear regressions with the price-dividend ratio as a predictive variable. They discover that their model is superior to ordinary linear regression for both cash and market-invested dividends. Additionally, they emphasize that it is extremely important to study predictability of stock returns and dividend growth rates jointly because there is a tight relationship between the predictive coefficients of returns and dividend growth rates and the persistence of the dividend yield.

As long as log-linearized present-value model relates price dividend ratio to the expected returns and expected dividend growth rates, there is an alternative methodology, which does not require complex estimation techniques. The idea comes from Lacerda and Santa-Clara (2010), who argue that one can adjust price-dividend ratio for variations in expected growth rates and use the adjusted ratio to forecast future returns. However, this approach requires assuming how market participants estimate future dividend growth rates. Lacerda and Santa-Clara (2010) use moving average of historical dividend growth rates as an estimate for future dividend growth. However, they reasonably stress the fact that the existence of better predictors is an open question and if such predictors are found, they will presumably lead to better estimates of expected returns. Then, they transform an initial dividend-price ratio as follows:

$$x_t = dp_t + \frac{\bar{g}_t}{1 - \bar{\rho}_t} \quad (1.1)$$

where  $dp_t$ - log dividend-price ratio;

$\bar{g}_t$ - historical averaged log dividend growth;

$$\bar{\rho}_t = \frac{1}{1 + \left(\frac{D}{P}\right)_t}$$

Finally, they use this new adjusted ratio as a predictive variable. The intuition behind the adjustment is to distinguish between changes in dividend-price ratio due to changes in expected dividend growth rate and due to changes in expected future returns. They find out that adjusted dividend yield explains more variation in future returns than ordinary dividend yield both in- and out-of-sample.

The fact that Lacerda and Santa-Clara reveal statistically significant predictive power even out-of-sample becomes even more important in the light of Goyal and Welch (2008), who strongly criticize any evidence of returns predictability. They argue that in the real world we cannot use information that is not available yet. Good in-sample performance of some valuation ratios and other predictors is not practically important. Even if the true model exists and it is known, the true coefficients are unknown and we have to adjust their estimates as new data become available. Hence, for practical purposes Goyal and Welch measure out-of-sample performance and find that all common predictors show very poor results. Goyal and Welch claim that it is still in-sample performance that indicates an

empirical fit of the model; however, it should be studied jointly with out-of-sample tests in order to assess applicability of in-sample results. Therefore, it became natural to evaluate promising predictors out-of-sample since the paper of Goyal and Welch.

### 3. Data

We obtain monthly data of discrete total returns and capital gains (price returns) on the value-weighted portfolio of all NYSE, Amex and Nasdaq stocks from January, 1926 to December, 2011. The data is provided by Center for Research in Security Prices (CRSP). Each month all values are recorded in the last trading day. For example, a value of capital gain on the index in April, 1990 corresponds to a percentage increase in price between 30<sup>th</sup> of March and 30<sup>th</sup> of April in 1990. Further we extract monthly price and dividend series from the returns. It is typical to avoid monthly data when studying stock returns and dividend growth rates predictability, because it was repeatedly documented that dividend series exhibit strong seasonality patterns, which can drive dividend growth predictability. *Exhibit 1* provides a convenient support for the presence of seasonality trends in monthly dividend growth data. Log dividend growth rates are grouped by month and mean is presented for each group.

*Exhibit 1. Average dividend growth rates for monthly data form February, 1926 to December, 2011, %*

Month	Jan	Feb	Mar	Apr	May	Jun
Average log dividend growth	-74.47	100.79	-36.94	-62.76	106.67	-41.96
Month	Jul	Aug	Sep	Oct	Nov	Dec
Average log dividend growth	-60.83	101.35	-47.43	-42.91	109.91	-46.82

There is an obvious regularity in means: dividends in February, May, August and November on average are more than twice as high as dividends in January, April, July and October. We further conduct more formal analysis in order to confirm seasonality. We estimate a simple linear regression model with dividend growth rate as a dependent variable and two dummy variables as covariates. First dummy variable takes a value of 1 in February, May, August and November and 0 otherwise. The second variable equals 1 in March, June, September and December and 0 otherwise. The coefficients of these variables can be interpreted as an effect



of the aforementioned months on dividend growth rate relative to January, April, July and October. The results are summarized in *Exhibit 2*.

*Exhibit 2. Estimation results for regression of dividend growth rates on season dummies for monthly data*

Dependent variable	Dividend growth rate
Independent variables	
Constant	-0.602*** (0.024)
D1 (February, May, August, and November)	1.649*** (0.034)
D2 (March, June, September, and December)	0.169*** (0.034)
$R^2$	0.732

*Standard errors are given in parentheses. \*\*\* denotes significance at the 1% level.*

Both dummy coefficients are positive and extremely significant confirming our expectations about seasonality patterns. Moreover,  $R^2$  of 0.732 indicates that such simple way of controlling for seasonality already explains 73% of variation in dividend growth rates for monthly data. We conclude that monthly model should not be used for further analysis as seasonality patterns can completely invalidate the results. Still one may be confused by periodicity of the pattern that equals to 3 months. The intuition behind this pattern is the following: it has become a common practice for majority of dividend paying firms to use a quarterly basis to pay dividends. Then, there is a second question. What is special about February, May, August and November relative to January, April, July and October? In our opinion, it is mainly driven by May and November. The former implies increase in consumption due to upcoming summer vacation, while the latter is associated with preparation for Christmas days, which are accompanied by an increased consumption as well. Therefore, a lot of firms choose February-May-August-November scheme of dividend payments in order to fulfill shareholders' liquidity needs.

In order to avoid seasonality pattern in dividends in the remainder of the thesis we work with annual model. However, annual data has to be constructed from monthly observations and there is an important issue to be considered: we need to take into account how dividends received within a particular year are reinvested. There are three common ways to approach to this question. The easiest one is to ignore reinvestment at all. That is, annual dividends are computed as a sum of

dividends during last 12 months. This approach has an important drawback: it ignores time value of money. Even though it misses important information, some researchers still stick to it arguing that it reduces some technical work, while delivering practically the same results. The two other approaches are cash-invested and market-invested dividends. In essence, these two are extreme cases of reinvestment strategies. In cash-invested dividends case, each month within a particular year dividends are reinvested in one-month T-bill. Alternatively, it means that all dividends are consumed implying that the risk-free rate is an appropriate discount factor for computation of future value of dividends. In turn, in market-invested dividends it is assumed that dividends are reinvested in the aggregate stock market and nothing is spent on consumption. Both reinvestment strategies have been studied widely in the dividend-growth and returns predictability literature. We believe that the precise methodology lies somewhere in between, but it should be closer to cash-invested dividends. There is a well-known anomaly in the U.S. tax laws: dividends are taxed at higher rates than capital gains. It led to a lot of papers written in the area of dividend payments (for example, Fama and French (2000)). In particular, some researchers investigate why firms pay dividends at all given this anomaly. We do not touch this question in our thesis, but keep in mind that anomaly is important for the choice between two reinvestment strategies.

Consider a typical investor, who does not have short-term liquidity needs. It could be either long-term individual investor, whose salary is large enough to cover his living expenses, or large financial institution (for example, pension fund), that holds a portfolio for long-term needs. Since the portfolio of such investor is not required to support consumption, it will likely have a bias towards firms that do not pay dividends, because it decreases the effective tax rate. Therefore, we believe that the majority of investors, who care about dividends, use paid dividends for consumption rather than for reinvestment. It corresponds to the case of cash-invested dividends and we believe that it is the most convenient approach to construction of annual dividend data. Thereafter we work with cash-invested dividends only. The data on one-month T-bill rates is obtained from Kenneth French's webpage.

We further construct annual series of dividend growth rates, returns and price-dividend ratio. Descriptive statistics are summarized in *Exhibit 3*.

*Exhibit 3. Summary statistics of dividend growth rates, returns and price-dividend ratio.*

Summary statistics	$\Delta d_t$	$r_t$	$pd_t$
Mean	0.043	0.090	3.356
Median	0.055	0.132	3.326
Standard Deviation	0.123	0.195	0.446
Skewness	-1.800	-0.989	0.456
Excess Kurtosis	8.102	1.184	-0.099
Minimum	-0.520	-0.556	2.352
Maximum	0.423	0.447	4.457
# of observations	86	86	87

We observe a substantial difference between mean and median values for both returns and dividend growth series. This difference is negative and driven by large negative shocks of the Great Depression and the recent financial crisis. Both these shocks push down the mean values, while have a negligible effect on the medians. Price-dividend ratio evolves slowly over time implying a relatively low standard deviation of 0.446 and tight bounds of 2.352 and 4.457.

It is worth mentioning that some researchers tend to adjust the sample period by exclusion of extreme outliers. For example, it is typical to exclude two aforementioned major shocks and the Second World War period ending up with a sample of 1946-2007. One may argue that most statistical procedures are calibrated for normally distributed data and exclusion of outliers, which practically cannot be observed in the case of normal distribution, is an attempt to reach normality. We do not share this opinion for at least three reasons.

First of all, the rule of determination whether an observation is outlier or not is very vague. Clearly, the advantage of the Great Depression and the 2008 financial crisis is their timing in the beginning and the end of the sample respectively making them easy to exclude. Still there are such events as 1973-1974 stock market crash after the collapse of the Bretton Woods system and the Dot Com bubble followed by a series of downturns in 2000-2002, which also led to negative returns of the decent scale as the Great Depression and the 2008 crisis; hence, deserve to be considered as outliers. However, they cannot be excluded as

they are located in the middle of the sample. In addition, there are years of economic booms with positive returns of great magnitude. The fact that severe shocks happen from time to time means that these events should not be treated as extreme. Instead, we should conclude that distribution of returns is far from Normal and we have to live with it.

Secondly, even if we are able to eliminate all outliers and normalize the kurtosis, there is a second feature, which is typically observed in financial time series: negative skewness. It implies substantial asymmetry reflected in a longer left tail of distribution and prevents data set from passing even a low-power Jarque-Bera test for normality. It is indeed the case for both returns and dividend growth series with the skewness of -0.989 and -1.8 respectively.

Finally, it is true that normality is a desirable feature of dataset for the most of statistical models. However, the Central Limit Theorem ensures that such major statistical tools as confidence intervals and hypothesis testing are asymptotically valid (i.e. as sample size tends to infinity) even when the data is not normal. Practically, it means that when the sample size is sufficiently large, we are not really concerned with the normality of data. However, we are still concerned with such issues as homogeneity of residuals, omitted variable bias and serial autocorrelation.

Taking into account the argument above, we conclude that the entire sample (1926-2011) should be included in the analysis. Therefore, we end up with 87 observations of price-dividend ratio and 86 observations of returns and dividend growth rate.

#### **4. Present value models**

Campbell and Shiller (1988) introduce a log-linearized present value model that approximates realized returns as a linear function of price-dividend ratio, lagged price dividend ratio and dividend growth rate. Linearization is a typical methodology in empirical research as it substantially simplifies technical part of analysis and allows employing wide variety of tools, which could be applied to linear models only. For example, ordinary OLS estimation is valid only when a dependent variable is a linear function of unknown parameters. Otherwise, it may generate severely biased estimates. Moreover, linear functions are attractive for

their convenient interpretation. If we are given a linear function of several variables, we can easily compute the change in the function value after a change in one of the variables by examination of the respective coefficient.

In our work we present two versions of present-value model. Since our goal is a broad comparison of methodology used by Binsbergen and Kojien (2010) and predictive regressions with adjusted price-dividend ratio, we first introduce a model of Binsbergen and Kojien. Throughout the paper we refer to it as a benchmark model.

#### 4.1 Benchmark model

Binsbergen and Kojien's model is based on the assumption that both expected returns and expected dividend growth rates are latent and follow an AR(1) process, though it can be extended to higher order of ARMA family. This assumption of persistency evolves from the findings of previous works of Fama and French (1988), Campbell and Cochrane (1999), and Pastor and Stambaugh (2006). The first-order autoregressive component is consistently found to be significant for both expected returns and expected dividend growth rates. Intuitively we should expect it as business cycle theories typically find a strong empirical support. Recessions and expansions follow each other, but overall economic conditions change very smoothly over time. It is natural to believe that expected returns and dividend growth rates are driven by current and expected future economic conditions. As the latter is highly persistent, we should not be surprised by persistency in expected returns and expected dividend growth.

In our work we use conventional notation for log return, log price-dividend ratio and log dividend growth rate:

$$r_{t+1} \equiv \log\left(\frac{P_{t+1} + D_{t+1}}{P_t}\right) \quad (4.1.1)$$

$$pd_t \equiv \log\left(\frac{P_t}{D_t}\right) \quad (4.1.2)$$

$$\Delta d_{t+1} \equiv \log\left(\frac{D_{t+1}}{D_t}\right) \quad (4.1.3)$$

Expected return and expected dividend growth rate denoted as  $\mu_t$  and  $g_t$  accordingly are assumed to follow AR(1)-process:

$$\mu_{t+1} = \delta_0 + \delta_1(\mu_t - \delta_0) + \varepsilon_{t+1}^\mu \quad (4.1.4)$$

$$g_{t+1} = \gamma_0 + \gamma_1(g_t - \gamma_0) + \varepsilon_{t+1}^g \quad (4.1.5)$$

where  $\mu_t = E_t[r_{t+1}]$ ,  $g_t = E_t[\Delta d_{t+1}]$ ,  $\delta_0$  – unconditional mean of expected returns (if  $|\delta_1| < 1$ ),  $\gamma_0$  – unconditional mean of expected dividend growth rates (if  $|\gamma_1| < 1$ ).

Realized dividend growth rate can be decomposed into the sum of expected growth rate and an orthogonal shock:

$$\Delta d_{t+1} = g_t + \varepsilon_{t+1}^D \quad (4.1.6)$$

We proceed to linearization of the model for log returns, departing from the following identity:

$$r_{t+1} = \log\left(1 + \frac{P_{t+1}}{D_{t+1}}\right) + \Delta d_{t+1} - pd_t \quad (4.1.7)$$

Using a Taylor expansion around the historical average of  $pd_t$  and ignoring all terms after the first order, we get:

$$\begin{aligned} r_{t+1} &= \log(1 + \exp(pd_{t+1})) + \Delta d_{t+1} - pd_t = \\ &= \log(1 + \exp(\overline{pd})) + \frac{\exp(\overline{pd})}{1 + \exp(\overline{pd})} (pd_{t+1} - \overline{pd}) + o(pd_{t+1} - \overline{pd}) + \\ &\quad + \Delta d_{t+1} - pd_t \cong \kappa + \rho pd_{t+1} + \Delta d_{t+1} - pd_t \end{aligned} \quad (4.1.8)$$

where  $\overline{pd} = \sum_{t=1}^T \frac{pd_t}{T}$ ,  $\kappa = \log(1 + \exp(\overline{pd})) - \rho \overline{pd}$  and  $\rho = \frac{\exp(\overline{pd})}{1 + \exp(\overline{pd})}$ .

Even though this linearization introduces an approximation error, it is consistent with the economic intuition. Higher dividend growth rates imply higher cash flows and, thus, positively affect returns. Increase in price leads to both increase in price-dividend ratio and decrease in expected returns which on average leads to reduction of realized next period return. Finally, higher current price means higher return in the same period.

Assuming that the result from (4.1.8) is an exact equality rather than an approximation, one can iterate it and get:

$$\begin{aligned} pd_t &= \kappa + \rho pd_{t+1} + \Delta d_{t+1} - r_{t+1} = \\ &= \kappa + \rho(\kappa + \rho pd_{t+2} + \Delta d_{t+2} - r_{t+2}) + \Delta d_{t+1} - r_{t+1} = \\ &= \kappa + \rho\kappa + \rho^2 pd_{t+2} + \Delta d_{t+1} - r_{t+1} + \rho(\Delta d_{t+2} - r_{t+2}) = \\ &= \kappa \sum_{i=0}^{\infty} \rho^i + \lim_{n \rightarrow \infty} \rho^n pd_{t+n} + \sum_{i=1}^{\infty} \rho^{i-1} (\Delta d_{t+i} - r_{t+i}) \end{aligned} \quad (4.1.9)$$

The next step is to use expectations conditional on information available at time  $t$ . It is natural to assume that  $\lim_{n \rightarrow \infty} \rho^n \cdot E_t(pd_{t+n}) = 0$ , since  $|\rho| < 1$  and variation of  $pd_t$  is likely to be limited in the long run. Then,

$$\begin{aligned} pd_t &= \frac{\kappa}{1-\rho} + \sum_{i=1}^{\infty} \rho^{i-1} (E_t[\Delta d_{t+i}] - E_t[r_{t+i}]) = \\ &= \frac{\kappa}{1-\rho} + \sum_{i=0}^{\infty} \rho^i (E_t[g_{t+i}] - E_t[\mu_{t+i}]) \end{aligned} \quad (4.1.10)$$

Using the properties of autoregressive processes one can show that:

$$E_t[\mu_{t+k}] = \delta_0 + \delta_1^k (\mu_t - \delta_0) \quad (4.1.11)$$

$$E_t[g_{t+k}] = \gamma_0 + \gamma_1^k (g_t - \gamma_0) \quad (4.1.12)$$

Thus,

$$\begin{aligned} pd_t &= \frac{\kappa}{1-\rho} + \sum_{i=0}^{\infty} \rho^i (\gamma_0 + \gamma_1^i (g_t - \gamma_0) - (\delta_0 + \delta_1^i (\mu_t - \delta_0))) = \\ &= \frac{\kappa}{1-\rho} + \frac{\gamma_0 - \delta_0}{1-\rho} + \frac{g_t - \gamma_0}{1-\rho\gamma_1} - \frac{\mu_t - \delta_0}{1-\rho\delta_1} = \\ &= A - B_1(\mu_t - \delta_0) + B_2(g_t - \gamma_0) \end{aligned} \quad (4.1.13)$$

Equation (4.1.13) illustrates that price-dividend ratio is only a noisy proxy for expected returns, as it is also affected by changes in expected dividend growth rate. Moreover, it shows that negative expected return and positive expected dividend growth rate shocks positively affect price dividend ratio, which is in line with economic intuition behind these variables. The magnitude of change in price-dividend ratio depends on the persistency coefficients:  $\delta_1$  and  $\gamma_1$ .

In their paper, Binsbergen and Kojien claim that their theoretical framework and further estimation procedure aggregate the whole history of price-dividend ratio and dividend growth rates to predict future returns and future dividend growth rates. They argue that for that reason their approach has better predictive power than standard predictive regressions have. However, in the light of Goyal and Welch (2008) critique of real time applicability of standard forecasting models, we are interested in checking of out-of-sample performance of Binsbergen and Kojien model. Apart from that, we also replicate their methodology in-sample for our data set. Detailed description of estimation procedure and related statistical issues are given in *Section 5.1.1*.

#### 4.2 Adjusted price-dividend ratio

In this section we propose an alternative methodology to study stock returns predictability. We develop our model that is also based on the present-value identity, but evolves from different assumptions than those used in Binsbergen and Koijen paper.

As was emphasized by equation (4.1.13), variation of price-dividend ratio reflects not only changes in expected returns, but also changes in expected future dividends. Therefore, we should not expect strong predictive power of regular price-dividend ratio in return forecasting regression. On the other hand, this equation gives rise to a new approach, which does not require complex estimation techniques and modeling of latent processes. This approach was first introduced by Lacerda and Santa-Clara (2010) and was also used by Golez (2012).

Since equation (4.1.13) links expected returns with price-dividend ratio and expected dividend growth rate, one can come up with reasonable estimates of future dividend growth and improve returns predictability. More precisely, our model relies on the following assumptions:

1. Latent expected returns follow an AR(1) process as in equation (4.1.4) and parameters generating the process are known to investors;
2. The entire process for dividend growth is unknown to investors, which means that they have to forecast it using, for example, historical data;
3. Investors price the market in accordance with their estimates of future dividend growth.

Linearization and iteration procedures from the previous section are not affected by new assumptions, and the following equation is still valid.

$$pd_t = \frac{\kappa}{1 - \rho} + \sum_{i=1}^{\infty} \rho^{i-1} (E_t[\Delta d_{t+i}] - E_t[r_{t+i}]) \quad (4.2.1)$$

Moreover, the part of the equation representing expected returns does not change as well, and we get:

$$pd_t = \frac{\kappa - \delta_0}{1 - \rho} - \frac{\mu_t - \delta_0}{1 - \rho \delta_1} + \sum_{i=1}^{\infty} \rho^{i-1} E_t[\Delta d_{t+i}] \quad (4.2.2)$$



This equation shows the relationship between log price-dividend ratio, one period expected return, and expected future dividend growth rates. If we slightly rearrange the terms, we get the following expression for the expected return:

$$\mu_t = \delta_0 + (1 - \rho\delta_1) \left( \frac{\kappa - \delta_0}{1 - \rho} + \sum_{i=1}^{\infty} \rho^{i-1} E_t[\Delta d_{t+i}] - pd_t \right) \quad (4.2.3)$$

From this equation we can see that fluctuations in the price-dividend ratio that are solely due to changing forecasts of dividend growth do not affect expectations of future returns. In other words, higher forecasts of dividend growth increase price-dividend ratio, but this increase is neutralized by the second term in parentheses. It is worth mentioning that if investors' forecasts of dividend growth do not change over time (i.e.  $\sum_{i=1}^{\infty} \rho^{i-1} E_t[\Delta d_{t+i}]$  is constant), then price-dividend ratio depends solely on the expected returns. In this case it eventually becomes a good predictor of future returns. However, since investors' forecasts of dividend growth are time-varying, price-dividend ratio alone cannot be a perfect predictor. Then, instead of using only price-dividend ratio to predict returns, we must also take into account all future expected dividend growth rates. Thus, we obtain the following return forecasting equation:

$$\begin{aligned} r_{t+1} &= \varphi + (1 - \rho\delta_1) \left( \sum_{i=1}^{\infty} \rho^{i-1} E_t[\Delta d_{t+i}] - pd_t \right) + v_{t+1}^r = \\ &= \varphi - (1 - \rho\delta_1) pd_t^{adj} + v_{t+1}^r \end{aligned} \quad (4.2.4)$$

We call  $\sum_{i=1}^{\infty} \rho^{i-1} E_t[\Delta d_{t+i}] - pd_t$  an adjusted price-dividend ratio, because it corrects regular price-dividend ratio for expected future dividend growth rates. Ideally, it must be a better predictor for future returns, as it is not affected by variation in future dividends. The exact form of adjustment can be revealed by our second assumption. That is, we have to specify how exactly investors estimate dividend growth rates.

For this purpose we consider three different alternatives, each of them relies on historical data. They are historical average forecast, ARMA(1,1)-process for realized dividend growth rates, and AR(1)-process for realized dividend growth rates.

### 4.2.1 Historical mean adjustment

The simplest way to estimate future values of time series is historical mean. During such estimation one implicitly assumes that the respective time series are well described by a random walk process. If this is the case, the historical mean forecast should outperform other complicated predictors. Indeed many empirical works show that it is true for many financial variables. Thus, it is tempting for investors to use this approach with dividend growth rates.

As Lacerda and Santa-Clara (2010) do in their paper, we assume that investors forecast all future dividend growth rates using moving historical average:

$$E_t[\Delta d_{t+i}] = \overline{g_{t,s}} = \sum_{j=t-s+1}^t \frac{\Delta d_j}{s}, \forall i \quad (4.2.1.1)$$

In this case the expression for expected future dividend growth rates can be simplified as follows:

$$\sum_{i=1}^{\infty} \rho^{i-1} E_t[\Delta d_{t+i}] = \frac{\overline{g_{t,s}}}{1-\rho} \quad (4.2.1.2)$$

Then we get the following expression for the adjusted price-dividend ratio:

$$pd_t^{adj} \equiv pd_t^{HM} = pd_t - \frac{\overline{g_{t,s}}}{1-\rho} \quad (4.2.1.3)$$

The choice of  $s$  is a sophisticated issue. Large values of  $s$  produce stable dividend growth estimates, which evolve slowly over time. In turn, low values may be more efficient in capturing current economic conditions since they are not affected by old data. However, they can also induce additional noise to forecasts, which is an undesirable feature. Lacerda and Santa-Clara (2010) propose  $s = 10$ , which roughly represents a full business cycle. In order to completely replicate their procedure, we use this value of  $s$ . However, we additionally consider  $s = 20$  and long-term average of dividend growth rates, which incorporates all available dividend growth data at the time when next period growth is forecasted. Given the assumption above, we can compute the last component from (4.8):

$$pd_t = \frac{\kappa - \delta_0}{1-\rho} - \frac{\mu_t - \delta_0}{1-\rho\delta_1} + \frac{\overline{g_{t,s}}}{1-\rho} \quad (4.2.1.2)$$

If we slightly rearrange the terms, we can also derive an expression for the expected returns:

$$\begin{aligned}\mu_t &= \delta_0 + (1 - \rho\delta_1) \left( \frac{\kappa - \delta_0}{1 - \rho} + \frac{\overline{g_{t,s}}}{1 - \rho} - pd_t \right) = \\ &= \delta_0 + (1 - \rho\delta_1) \left( \frac{\kappa - \delta_0}{1 - \rho} - pd_t^{HM} \right)\end{aligned}\quad (4.2.1.3)$$

#### 4.2.2 ARMA adjustment

Since Lacerda and Santa-Clara leaves an open question of other models for dividend growth rates, it is natural to propose another model, which is structurally equivalent to the model of Binsbergen and Koijen for cash-invested dividends. In other words, what is the benchmark model's implied adjustment for price-dividend ratio?

As was mentioned before, in state-space setup of Binsbergen and Koijen expected dividend growth follows AR(1)-process, while realized dividend growth is decomposed to its mean and orthogonal shock. More formally,

$$g_{t+1} = \gamma_0 + \gamma_1(g_t - \gamma_0) + \varepsilon_{t+1}^g \quad (4.2.2.1)$$

$$\Delta d_{t+1} = g_t + \varepsilon_{t+1}^D \quad (4.2.2.2)$$

where  $g_t = E_t[\Delta d_{t+1}]$ .

Cochrane (2008) conducts a comprehensive analysis of state-space models and their observed representations. In particular, he shows that AR(1)-process for expected dividend growth rates implies ARMA(1,1)-model for realized dividend growth with the same constant and AR(1) component. That is,

$$\Delta d_{t+1} = \gamma_0 + \gamma_1(\Delta d_t - \gamma_0) + (\beta - \gamma_1)\eta_t + \eta_{t+1} \quad (4.2.2.3)$$

where  $\eta_t$ — new error term (not equal to  $\varepsilon_t^D$ ).

Though an expression for  $\beta$  can be derived analytically, it is quite complex and does not exhibit any intuitive interpretation; therefore, we leave it as an unknown parameter, that is sufficient for estimation purposes.

If investors forecast future dividend growth according to this model, then these forecasts are as follows:

$$f_{t,i} = E_t[\Delta d_{t+i}] = \gamma_0 + \gamma_1^i(\Delta d_t - \gamma_0) + \gamma_1^{i-1}(\beta - \gamma_1)\eta_t \quad (4.2.2.4)$$

The initial expression for  $pd_t$  (4.1.13) can be simplified again:

$$pd_t = \frac{\kappa + \gamma_0 - \delta_0}{1 - \rho} - \frac{\mu_t - \delta_0}{1 - \rho\delta_1} + \frac{\gamma_1(\Delta d_t - \gamma_0) + (\beta - \gamma_1)\eta_t}{1 - \rho\gamma_1} \quad (4.2.2.5)$$

$$\begin{aligned} \mu_t &= \delta_0 + (1 - \rho\delta_1) \left( \frac{\kappa - \delta_0}{1 - \rho} + \frac{\gamma_0}{1 - \rho} + \frac{\gamma_1(\Delta d_t - \gamma_0) + (\beta - \gamma_1)\eta_t}{1 - \rho\gamma_1} - pd_t \right) = \\ &= \delta_0 + (1 - \rho\delta_1) \left( \frac{\kappa - \delta_0}{1 - \rho} - pd_t^{ARMA} \right) \end{aligned} \quad (4.2.2.6)$$

#### 4.2.3 AR adjustment

In our opinion, an assumption that investors employ ARMA(1,1)-model can be a bit artificial. Nevertheless, we still study it due to structural equivalence to the benchmark model. However, we also consider an AR(1)-process for realized dividend growth, because we believe that a simple process structure can be better assumption for the aggregate market. The process is specified as follows:

$$\Delta d_{t+1} = \gamma_0 + \gamma_1(\Delta d_t - \gamma_0) + \eta_{t+1} \quad (4.2.3.1)$$

In this setup,

$$f_{t,i}^s = E_t[\Delta d_{t+i}] = \gamma_0 + \gamma_1^i(\Delta d_t - \gamma_0) \quad (4.2.3.2)$$

$$pd_t = \frac{\kappa + \gamma_0 - \delta_0}{1 - \rho} - \frac{\mu_t - \delta_0}{1 - \rho\delta_1} + \frac{\gamma_1(\Delta d_t - \gamma_0)}{1 - \rho\gamma_1} \quad (4.2.3.3)$$

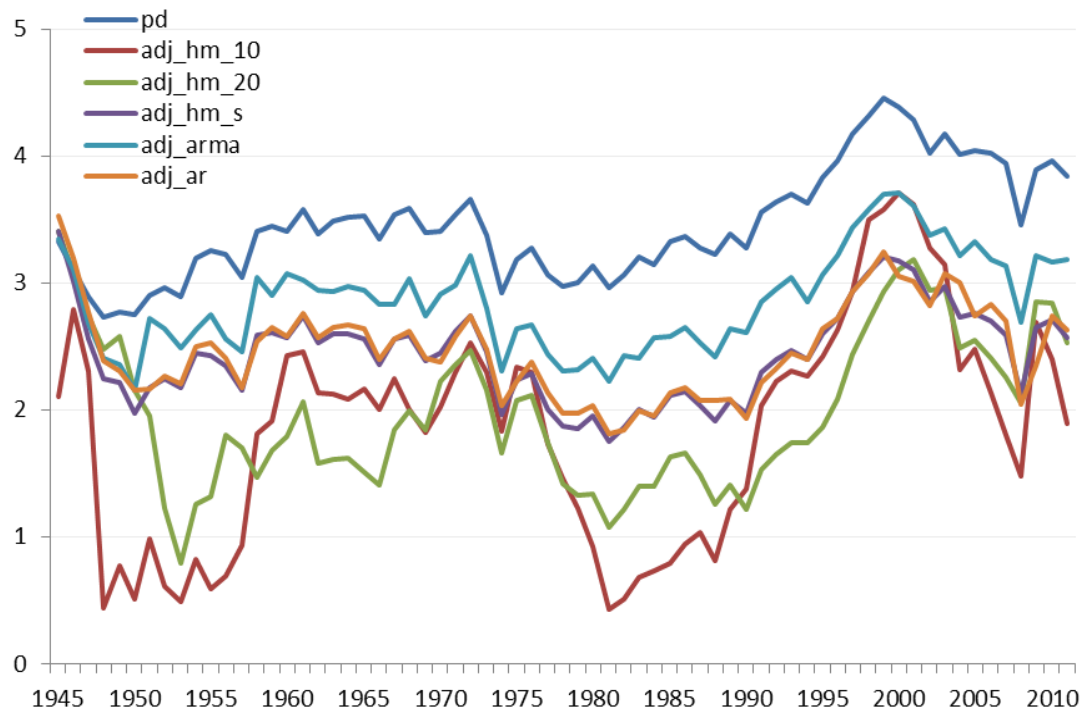
$$\begin{aligned} \mu_t &= \delta_0 + (1 - \rho\delta_1) \left( \frac{\kappa - \delta_0}{1 - \rho} + \frac{\gamma_0}{1 - \rho} + \frac{\gamma_1(\Delta d_t - \gamma_0)}{1 - \rho\gamma_1} - pd_t \right) = \\ &= \delta_0 + (1 - \rho\delta_1) \left( \frac{\kappa - \delta_0}{1 - \rho} - pd_t^{AR} \right) \end{aligned} \quad (4.2.3.4)$$

#### 4.2.4 How different the adjusted price-dividend ratios are?

Before we proceed to formal description of empirical methodology, we visually evaluate the difference in behavior of price-dividend ratio and its adjusted counterparts over 1945-2011 period<sup>1</sup>. All ratios are presented in *Figure 1*.

<sup>1</sup> We do not present it for the entire sample, because some adjusted ratios require 20 years of dividend growth data to be constructed. It is discussed more in depth in Section 5.1.3

*Figure 1. Evolution of adjusted and regular price-dividend ratios for the period 1945-2011*



*pd* – regular price-dividend ratio,

*adj\_hm\_10* – historical mean adjustment with 10 years of dividend growth data,

*adj\_hm\_20* – historical mean adjustment with 20 years of dividend growth data,

*adj\_hm\_s* – long-term historical mean adjustment with dividend growth data from the beginning of the sample,

*adj\_arma* – ARMA adjustment,

*adj\_ar* – AR adjustment.

We observe that all adjusted ratios are highly correlated with the regular price-dividend ratio and with each other. However, there are some years when adjusted and unadjusted ratios move in different directions. Theoretically, these should be the years, which will increase the predictive power of price-dividend ratio through the adjustment for expected dividend growth rates. It is worth mentioning that the historical mean adjustment with 10 years of dividend growth rates used for construction seems to be the most volatile and the least correlated with the regular price-dividend ratio. It is driven by very unstable forecast of dividend growth rate relative to other adjustments.

## 5. Empirical framework

In this section we present our methodology and introduce our approach to comparison of different models. We start with in-sample estimation, where we test

how different models fit the observed data over the entire sample period. Further we proceed to out-of-sample estimation, where we test if the models can be useful for investors in real time. We conclude with the construction of simple trading strategies based on the different models to check if they offer an opportunity to outperform the market in a consistent way.

### 5.1 In-sample estimation

The estimation processes for the benchmark model and the predictive regressions based on the adjusted-price dividend ratios do not have much in common. Further we describe them separately and discuss how the results should be interpreted and compared.

#### 5.1.1 In-sample estimation of the benchmark model

First of all, we de-mean both latent variables for convenience of notation.

$$\hat{\mu}_t = \mu_t - \delta_0,$$

$$\hat{g}_t = g_t - \gamma_0.$$

We have two simplified transition equations for latent variables:

$$\hat{\mu}_{t+1} = \delta_1 \hat{\mu}_t + \varepsilon_{t+1}^\mu,$$

$$\hat{g}_{t+1} = \gamma_1 \hat{g}_t + \varepsilon_{t+1}^g,$$

And two measurement equations, which link the observed variables to the underlying latent structure:

$$\Delta d_{t+1} = \gamma_0 + \hat{g}_t + \varepsilon_{t+1}^D,$$

$$pd_t = A - B_1 \hat{\mu}_t + B_2 \hat{g}_t.$$

We can simplify this system, because the last equation does not contain an error term. If we substitute it to the equation for the de-meaned expected returns, we decrease the number of transition equations and our final system consists of one transition equation and two measurement equations:

$$\hat{g}_{t+1} = \gamma_1 \hat{g}_t + \varepsilon_{t+1}^g,$$

$$\Delta d_{t+1} = \gamma_0 + \hat{g}_t + \varepsilon_{t+1}^D,$$

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

Additionally, we assume that a vector of the error terms has multivariate normal distribution with zero mean and the following covariance matrix:

$$\Sigma = \text{cov}(\varepsilon_{t+1}^D, \varepsilon_{t+1}^g, \varepsilon_{t+1}^\mu) = \begin{pmatrix} \sigma_D^2 & \sigma_D \sigma_g \rho_{Dg} & \sigma_D \sigma_\mu \rho_{D\mu} \\ \sigma_D \sigma_g \rho_{Dg} & \sigma_g^2 & \sigma_g \sigma_\mu \rho_{g\mu} \\ \sigma_D \sigma_\mu \rho_{D\mu} & \sigma_g \sigma_\mu \rho_{g\mu} & \sigma_\mu^2 \end{pmatrix}$$

The distribution is also assumed to be stationary over time.

Given the assumptions above the benchmark model has 10 unknown parameters to be estimated:

$$\Theta = (\gamma_0, \delta_0, \gamma_1, \delta_1, \sigma_D, \sigma_g, \sigma_\mu, \rho_{Dg}, \rho_{D\mu}, \rho_{g\mu})$$

The parameters are estimated by the conditional maximum likelihood estimator (MLE), while the likelihood of the model is constructed using a Kalman filter. The detailed description of the Kalman filter and a closed-form of the likelihood function are provided in *Appendix A*.

Due to a complex form of the likelihood function and potential existence of several local maximums, its global maximization becomes a separate serious problem. There is a well-known shortcoming of classical optimization methods (Quasi-Newton, Gradient descent): they are designed to find a local rather than global maximum. If the local maximum point is reached, the process is stuck there and one cannot identify if the type of the point is global maximum. In other words, these methods are sensitive to the initial guess of parameters. Nevertheless, in MLE we are interested in global maximization, and there are stochastic algorithms, which overcome the aforementioned problem. In the thesis we maximize likelihood using simulated annealing algorithm as Binsbergen and Koijen (2010) do. Relevant background information on simulated annealing is summarized in *Appendix B*. *Appendix C* is a programming code in MATLAB that was used for implementation of the simulated annealing algorithm.

After the likelihood function is maximized, the output contains not only a vector of estimated parameters, but also filtered expected returns and expected dividend growth rates, which are interpreted as the most likely values of the true expected returns and expected dividend growth rates given our assumptions. Further we use the filtered series for expected returns ( $\mu_t^F$ ) to compute  $R^2$  value:

$$R^2 = 1 - \frac{\text{var}(r_{t+1} - \mu_t^F)}{\text{var}(r_t)}$$

The important thing is that the interpretation of  $R^2$  value is the same as it is in the classical linear regression model: it shows the amount of variation in realized returns captured by the filtered series for expected returns. It equals 1 only if all forecasts are perfect and the closer it is to 1, the less variance is attributed to residuals.

### 5.1.2 In-sample estimation of predictive regressions based on adjusted price-dividend ratio

In order to evaluate our methodology, we first estimate a predictive regression for the simple price-dividend ratio:

$$r_{t+1} = \alpha^{pd} + \beta^{pd} * pd_t + \varepsilon_{t+1}^{pd}$$

We estimate this regression with OLS and compute t-statistic for  $\beta^{pd}$  and  $R^2$  value.

For the adjusted price-dividend ratios we construct a series of dividend growth rate forecasts using moving average of length 10 or 20 and long-term average from the beginning of the sample for historical mean adjustment. In turn, AR(1) and ARMA(1,1)-models for dividend growth rates are estimated for AR and ARMA adjustments respectively. At each point of time we use for estimation only data available by that moment in order to make dividend growth forecasts by the market realistic. AR(1)-model is estimated with OLS, while ARMA(1,1) estimation employs Non-linear Least Squares (NLS).

When the series of dividend growth forecasts are constructed, we are able to compute the adjusted-price dividend ratios for all three adjustments. Finally, we estimate the following predictive regressions for returns with OLS:

$$\begin{aligned} r_{t+1} &= \alpha^{HM} + \beta^{HM} * pd_t^{HM} + \varepsilon_{t+1}^{HM} \\ r_{t+1} &= \alpha^{ARMA} + \beta^{ARMA} * pd_t^{ARMA} + \varepsilon_{t+1}^{ARMA} \\ r_{t+1} &= \alpha^{AR} + \beta^{AR} * pd_t^{AR} + \varepsilon_{t+1}^{AR} \end{aligned}$$

We report t-statistics for the estimated effects of the adjusted price-dividend ratios and  $R^2$  values for each regression. We additionally conduct White's test for heteroscedasticity to determine if standard errors and consequently t-statistics need to be corrected.



### 5.1.3 Comparison of the models

We compare different models on the basis of  $R^2$  values, because it measures an amount of variation in realized returns explained by predictive variables in case of adjusted and usual price-dividend ratios and by filtered series of expected returns in case of the benchmark model. It is a consistent way of comparison only if  $R^2$  values of all models are computed for the same time period. That is, we need to choose the starting point for estimation. It should be noted that we cannot start from the beginning of the sample (i.e. 1926), because adjusted-price dividend ratio series require some time to construct the first element. For historical mean adjustment it is equivalent to the number of dividend growth rates used to compute the moving average, so it could be as high as 20 years. In turn, AR and ARMA adjustments appear to be very unstable in the first years due to low number of observations, but Figure 2 and Figure 3 suggest that after the first 20 years they are already fairly stable.

Figure 2. Stabilization of AR(1)-model for realized dividend growth rates

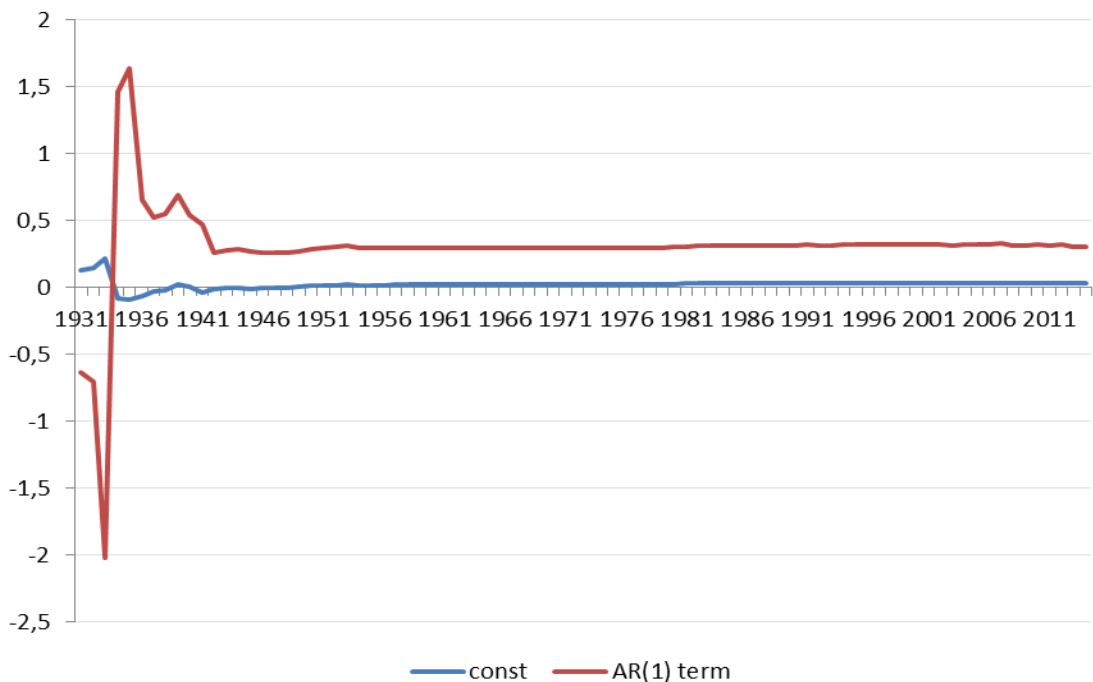
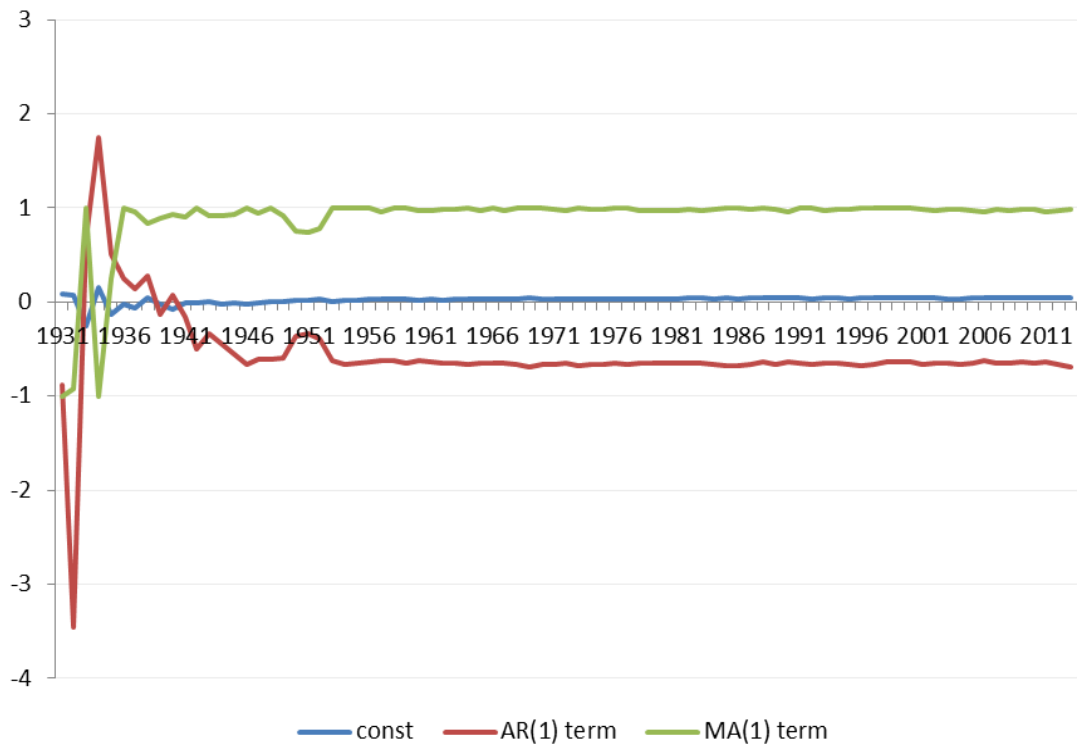


Figure 3. Stabilization of ARMA(1,1) model for realized dividend growth rates



We conclude that 20 years from the beginning of the sample is the most appropriate point of time to be used as the beginning of estimation. Thus, all predictive regressions are estimated for the period from 1945 to 2011. Binsbergen and Kojien model is estimated for the entire sample in order to account for dividend growth rates information before 1945, which is also incorporated into adjusted price-dividend ratios. However,  $R^2$  value for Binsbergen and Kojien model is computed for the filtered series of expected returns from 1945 to 2011 to enhance the comparison with predictive regressions.

### 5.2 Out-of-sample estimation

The importance of out-of-sample estimation cannot be overestimated. We employ the methodology introduced by Goyal and Welch (2008) to understand if any of the proposed predictive models can be used in real-time. Can the models be valuable to investors? And if they can, is there any consistent pattern, which does not depend on the choice of sample period?

In order to answer these questions, we estimate all models using the same methodology described in the previous section. However, the models are estimated separately at each point of time using only data that is available up to this moment. Further, the next year return is forecasted based on the estimated

model and forecast error is computed as the difference between the realized return and its prediction. These errors are employed to construct out-of-sample  $R^2$  series:

$$R_{OOS}^2 = 1 - \frac{MSE_R}{MSE_M}$$

where  $MSE_R$  is the mean squared error from the predictive model,  $MSE_M$  is the mean squared error from historical mean.

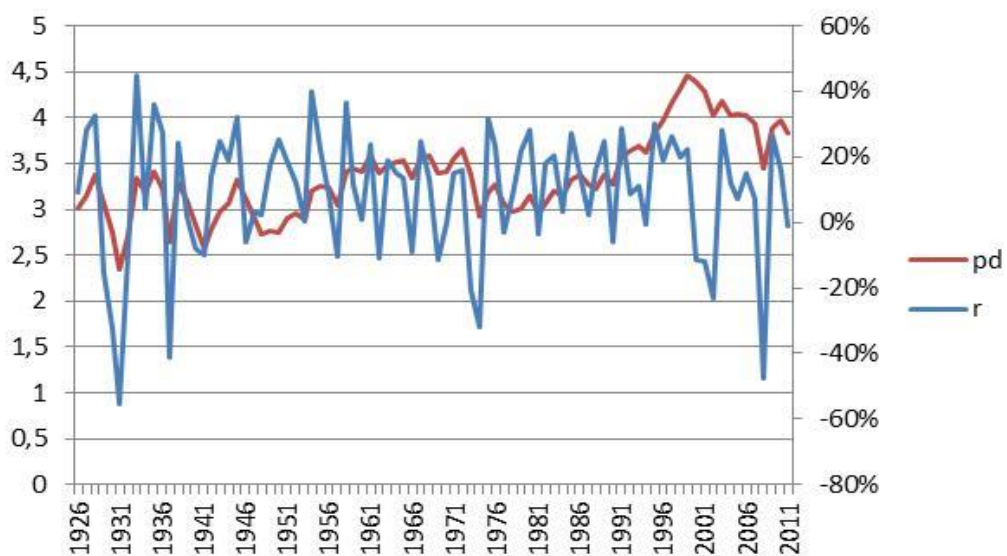
When  $R_{OOS}^2$  is greater than 0, the predictive model provides better forecasts than a simple historical mean forecast, as it generates smaller errors on average. Consequently, negative  $R_{OOS}^2$  implies that the model is unable to outperform the historical mean.

In order to ensure that  $R_{OOS}^2$  is not driven by the choice of the error aggregation starting point, we consider 4 different time frames beginning accumulation in 1948, 1963, 1978 and 1993. By doing so, we assess the robustness of the considered models.

### 5.3 Economic significance

When one studies stock returns predictability, he should understand in advance that the predictive power of any model that is based on price-dividend ratio is limited. It is sufficient to compare volatility of price-dividend ratio to volatility of realized returns.

Figure 4. Price dividend ratio and log returns, 1926-2011



Even if we strongly believe in the rational market pricing and price-dividend ratio precisely reflects expectations of future returns and dividends, there is much more behind the actual realized returns. They are driven by a combination of complex macroeconomic factors and price-dividend ratio fails to capture the effects of future shocks. Therefore, we should not expect high  $R^2$  values both in- and out-of-sample. One model may be better than another according to  $R^2$  criteria, but the difference is likely to be small and difficult to interpret.

In this section we present a simple trading strategy to show that even a small difference between the models in terms of  $R^2$  values results in a significant value creation for investors. Consider an investor with a typical quadratic expected utility function:

$$E(U) = \mu_P - \frac{\chi}{2} \sigma_P^2,$$

Where  $\mu_P$  – expected portfolio return,  $\sigma_P^2$  – portfolio variance,  $\chi$  –risk-aversion coefficient, which is assumed to be equal to 3 as proposed, for example, by Golez (2012).

We assume a simple world with only two investment opportunities for the investor: risk-free rate (30 day T-bill) and market portfolio. Moreover, we assume that the investor forecasts the next period market portfolio return based on one of the predictive models. He also uses long-term historical market volatility as an input to his trading strategy. Given the investment opportunity set and the forecast of market return, the expected utility function can be presented as follows:

$$E(U) = r_f + \omega(\hat{\mu}_M - r_f) - \frac{\chi}{2} \omega^2 \hat{\sigma}_M^2,$$

Where  $\hat{\mu}_M$  – investor's prediction of market return,  $\hat{\sigma}_M^2$  – estimated market variance,  $\omega$  – portfolio weight in the market,  $r_f$  –risk-free rate.

The investor will maximize his expected utility with the following weight in the market:

$$\omega^* = \frac{\hat{\mu}_M - r_f}{\chi \hat{\sigma}_M^2}$$

The investor repeats the procedure each year and rebalances his portfolio. We obtain the time series of the realized portfolio returns:

$$R_{t+1} = r_{f,t+1} + \omega_t(r_{t+1} - r_{f,t+1})$$

We eventually compute ex-post Sharpe ratio for the strategy and Certainty Equivalent return (CE):

$$SR = \frac{\bar{R}_{p,e}}{\hat{\sigma}_{p,e}},$$

$$CE = \bar{R}_p - \frac{\chi}{2} \hat{\sigma}_p^2,$$

Where  $\bar{R}_p$  and  $\hat{\sigma}_p^2$  are the mean and variance of portfolio return and the subscript  $e$  corresponds to the returns in excess of risk-free rate. In essence, ex-post Sharpe ratio reflects overall feasibility of trading strategy, while CE shows an average risk-free return which would be equivalent to investor's strategy in terms of his utility function.

Sometimes the models predict the return that is below the risk-free rate. In this case, the investor takes a short position in the market portfolio according to the strategy. Clearly, it is non-realistic as the expected return cannot fall below the risk-free rate, otherwise nobody will ever hold a stock. Therefore, we additionally consider a constrained trading strategy, where the investor invests 100% in the risk-free rate and 0% in the market if the model predicts the market return below the risk-free rate.

Finally, we compare ex-post Sharpe ratios and CE for different predictive models and the market. We consider different strategy initiation times (1948, 1963, 1978 and 1993) for robustness check. The economic significance of the predictive models is expected to be the same relative to each other as their out-of-sample performance; however, the actual benefits of trading strategies are expected to become clear and observable.

## 6. Estimation Results

### 6.1 In-Sample estimation results

In this section we present our ins-sample empirical results and discuss them. We start with estimation results for the benchmark model, then move to the predictive regressions and eventually compare the models with  $R^2$  criteria.

### 6.1.1 Benchmark Model

*Exhibit 4* presents the results of the maximum-likelihood estimation of Binsbergen and Koijen model for the period 1926-2011, where  $R^2$  values are computed using 1945-2011 sub-period.

*Exhibit 4 Maximum-likelihood estimation results of Binsbergen and Koijen model*

Maximum-likelihood estimates			
$\delta_0$	0.072	$\gamma_0$	0.043
$\delta_1$	0.938	$\gamma_1$	0.249
$\sigma_\mu$	0.014	$\sigma_g$	0.117
$\rho_{\mu g}$	0.168	$\sigma_D$	0.001
$\rho_{\mu D}$	-0.481		
Implied present-value model parameters			
A	3.507	$\rho$	0.966
$B_1$	10.734	$B_2$	1.317
Maximized Likelihood	532.68		
$R^2$ values			
$R^2$ (Returns)	10.0%	$R^2$ (Div)	14.8%

The unconditional expected log return is estimated to be 7.2% and the unconditional expected log dividend growth rate equals to 4.3%. Both numbers fall below the estimates of Binsbergen and Koijen which are 9% and 6.2% for expected return and dividend growth respectively. In our opinion, this discrepancy is driven by an extended sample period, which contains two major negative shocks. When it comes to other parameters, our estimation falls in line with the results of Binsbergen and Koijen. Expected returns show very strong persistency with the coefficient of 0.938, whereas expected dividend growth appears to be less persistent with the coefficient of 0.249. Consistently, expected dividend growth rate shocks are more volatile than those of the expected return. Filtered series of the expected returns explains 10% of variation in the realized returns, which is further compared to  $R^2$  value of the predictive regressions. We also compute  $R^2$  for the dividend growth series as 14.8%, which is not important for the main discussion, but again consistent with the paper of Binsbergen and Koijen.

### 6.1.2 Predictive regressions

We summarize results of estimation for all predictive regressions in *Exhibit 5*.

*Exhibit 5. In-sample estimation results for different predictive regressions*

Independent Variable: $r_{t+1}$						
Predictor	(1)	(2)	(3)	(4)	(5)	(6)
Constant	0.521 (3.251)	0.369 (3.948)	0.574 (3.807)	0.508 (4.017)	0.526 (4.121)	0.542 (3.675)
$pd_t$	-0.123 (-2.659)					
$pd_t^{HM10}$		-0.121 (-2.962)				
$pd_t^{HM20}$			-0.182 (-3.182)			
$pd_t^{HMLT}$				-0.168 (-3.278)		
$pd_t^{AR}$					-0.173 (-3.389)	
$pd_t^{ARMA}$						-0.155 (-3.035)
$R^2$	9.94%	12.06%	13.65%	14.38%	15.22%	12.58%

*t*-statistics are given in parentheses.

First of all, we should note that the estimated coefficients of all predictors including usual price-dividend ratio have correct negative signs and always statistically significant at the 1% level. However, usual price-dividend ratio explains only 9.9% of variation in the next period log return, which is the lowest  $R^2$  value among all considered predictors. It can be considered as the first empirical evidence of importance of price-dividend ratio adjustments. The historical mean adjustment appears to be slightly sensitive to the length of rolling window used for forecasting of next period dividends contradicting with Lacerda and Santa-Clara paper, who claim that the results almost do not change when they use long-term historical mean instead of 10 years.  $R^2$  equals to 14.3% for the long-term historical mean adjustment, whereas it is 12% only for the 10 years adjustment. In turn,  $R^2$  for 20 years adjustment lies in between and equals to 13.7%. AR price-dividend ratio adjustment shows the best in-sample performance

among all predictors explaining 15.2% of variation in the next period log returns, which is more than 50% above  $R^2$  value for the simple price-dividend ratio. Surprisingly, ARMA adjustment appears to be only in the middle with  $R^2$  value of 12.6% outperforming only regular price-dividend ratio and 10 years historical mean adjustment. The likely explanation is extra noise generated by the inclusion of moving average term to the model for realized dividend growth rates. That is, the adjustment still neutralizes variations in future dividend growth expectations to some extent. However, it also adds more undesirable variability to the price-dividend ratio due to increased flexibility of dividend growth model, which actually deteriorates the predictability of returns relative to more stable AR and long-term historical mean adjustments.

### *6.1.3 Comparison of benchmark model with predictive regressions*

As was discovered earlier, the model of Binsbergen and Kojien achieves  $R^2$  of 10% for returns, which is practically the same as  $R^2$  in the predictive regression with simple price-dividend ratio and clearly below explanatory power of all adjusted ratios. Does it mean that the latent variables approach adds nothing to the predictive power of price-dividend ratio? The answer to this question is two-fold. As the main goal of our thesis is comparison of different models from perspective of returns predictability, it is true that predictive regressions for adjusted ratios outperform the model of Binsbergen and Kojien in-sample. However, we should not forget that the latter deals not only with predictability of returns, but also with dividend growth rates. In essence, the model of Binsbergen and Kojien allocates the predictability between returns and dividend growth rates, whereas adjusted ratios approach simply uses exogenously specified model for dividend growth rates without verification of its relevance. In other words, adjusted price-dividend ratio methodology achieves better results in returns predictability, because the predictability of dividend growth rates is fixed by assumption.

Overall we conclude that all models show promising results from the in-sample perspective and we proceed to evaluation of their out-of-sample performance.

### *6.2 Out-of-sample estimation results*

In this section we present out-of-sample estimation results for the benchmark model, regular price-dividend ratio and all adjusted ratios. For the historical mean



adjustment we now consider only long-term adjustment since it outperformed both short-term adjustments in-sample. Following our methodology we start aggregation of forecast errors at four different years: 1948, 1963, 1978 and 1993. *Exhibit 6* summarizes estimation results for all considered predictors.

*Exhibit 6. Out-of-sample  $R^2$  values in 2011 for different predictors with accumulation of errors starting in 1948, 1963, 1978 and 1993*

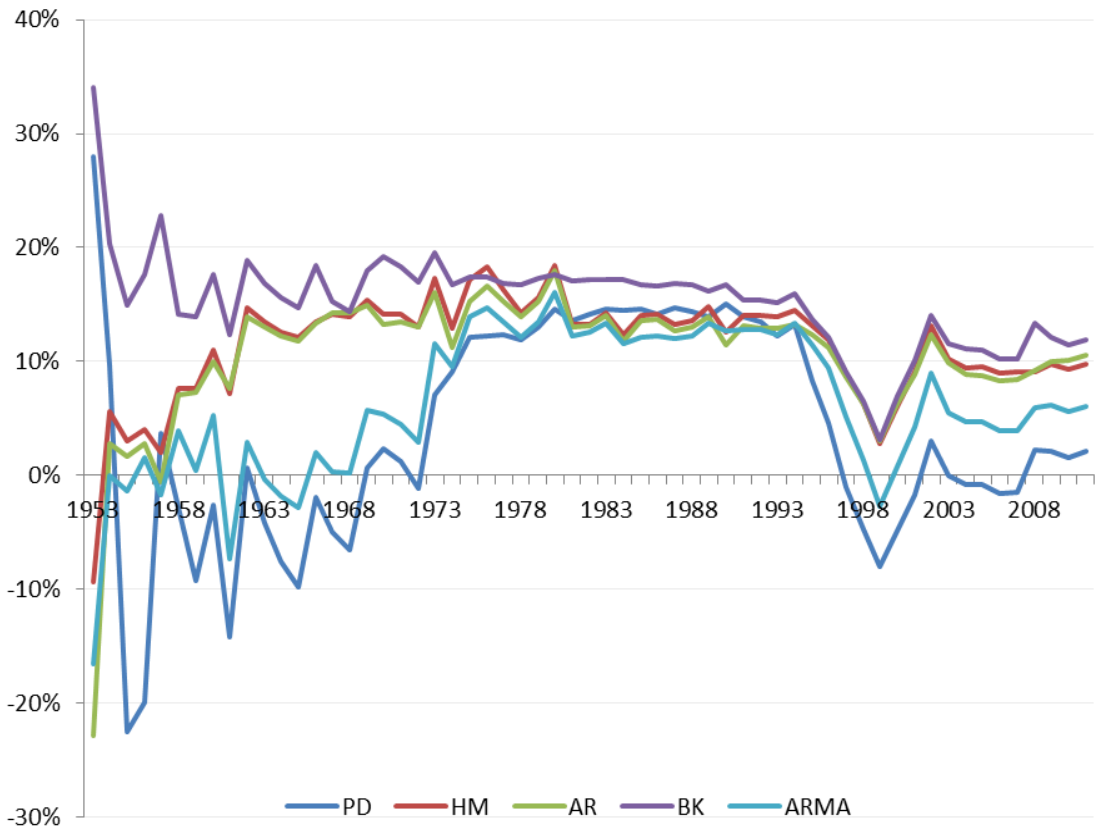
Aggregation period	PD	HM	AR	ARMA	BK
1948-2011	2.09%	9.77%	10.53%	6.03%	11.95%
1963-2011	2.45%	8.56%	9.71%	6.79%	10.24%
1978-2011	-7.29%	3.80%	6.24%	-0.71%	7.43%
1993-2011	-14.38%	3.67%	7.02%	-3.74%	6.95%

First of all, regular price-dividend ratio consistently shows the worst out-of-sample performance among all predictors. For the last two periods  $R^2$  values are even negative implying that it is not able to beat the historical mean forecasts. Among adjusted price-dividend ratios we observe the same pattern as it is for in-sample estimation. Both historical mean and AR adjustments always deliver positive  $R^2$  values, but AR adjustment always appears to be slightly better with particularly high  $R^2$  of 10.53% for the longest period. In turn, ARMA adjustment confirms our expectations about additional noise that it brings to returns predictability. It still performs better than regular price-dividend ratio, but  $R^2$  is negative for the last two periods, whereas historical mean and AR adjustments still able to predict returns in real time. Quite surprisingly, out-of-sample estimation of the benchmark model reveals that it is superior to all adjusted price-dividend ratios. It delivers an outstanding  $R^2$  of 11.95% for the longest period and keeps the leadership in 1963-2011 and 1978-2011 periods. However, in 1993-2011 it is outperformed by AR adjustment, but still has  $R^2$  of 6.95%, which is fairly large for this generally unfavorable period.

Out-of-sample  $R^2$  values in 2011 for different periods do not provide us with full understanding of long-term performance. In order to assess it, we construct full series of  $R^2$  values for each period. *Figure 5*, *Figure 6*, *Figure 7* and *Figure 8* present evolution of  $R^2$  values over time for 1948-2011, 1963-2011, 1978-2011 and 1993-2011 respectively. In each case we do not show first 5 years of the

period on the graph, because  $R^2$  values in the beginning of accumulation can be very unstable (i.e. values can exceed 100), thus it is difficult to see further development on the graph. In 5 years,  $R^2$  typically achieves its normal range.

*Figure 5. Evolution of OOS  $R^2$  for different predictors (accumulation starts in 1948)*

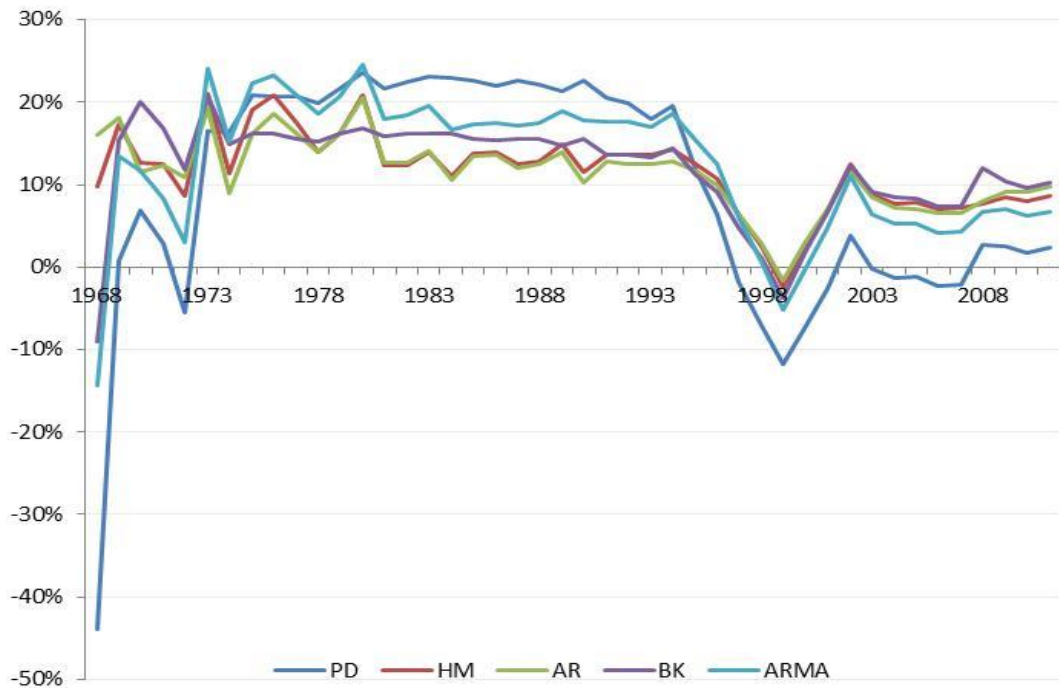


*PD, HM, AR, BK and ARMA denote regular price-dividend ratio, long-term historical mean adjustment, AR adjustment, benchmark (Binsbergen and Koijen) model, and ARMA adjustment accordingly.*

The most interesting thing for the longest period is a speed of  $R^2$  stabilization for the benchmark model. It is clearly the only predictor that stays at the 16% level for the major part of time. It is also negatively affected to the same extent as other predictors in 1999; however, it further recovers faster and reaches the value of 11.95%, which was documented earlier. Regular price-dividend ratio performs poorly in the first years, but it also has a period of good performance in the 70's, when it reaches other predictors and stays at the same level until 1994. It is further pushed downwards to negative values and does not show any signs of fast recovery.  $R^2$  values of AR and historical mean adjustments closely follow each other and show fairly stable performance never falling below zero after the first

years of the period. Finally,  $R^2$  of ARMA adjustments consistently stays below other adjusted ratios and it is negative in 1999.

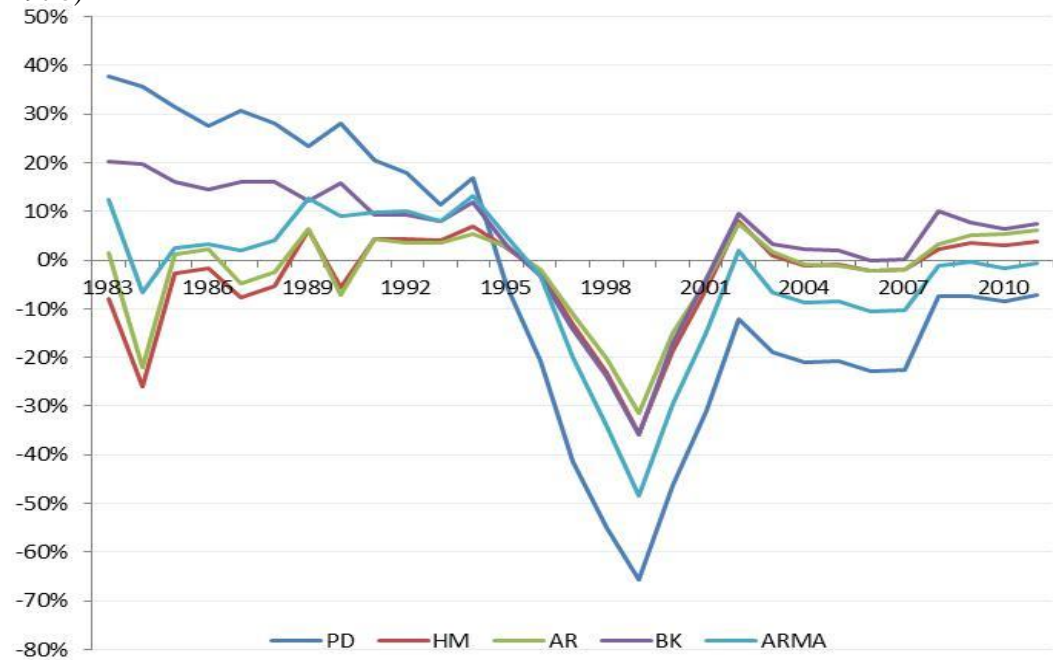
*Figure 6. Evolution of OOS  $R^2$  for different predictors (accumulation starts in 1963)*



*PD, HM, AR, BK and ARMA denote regular price-dividend ratio, long-term historical mean adjustment, AR adjustment, benchmark (Binsbergen and Koijen) model, and ARMA adjustment accordingly.*

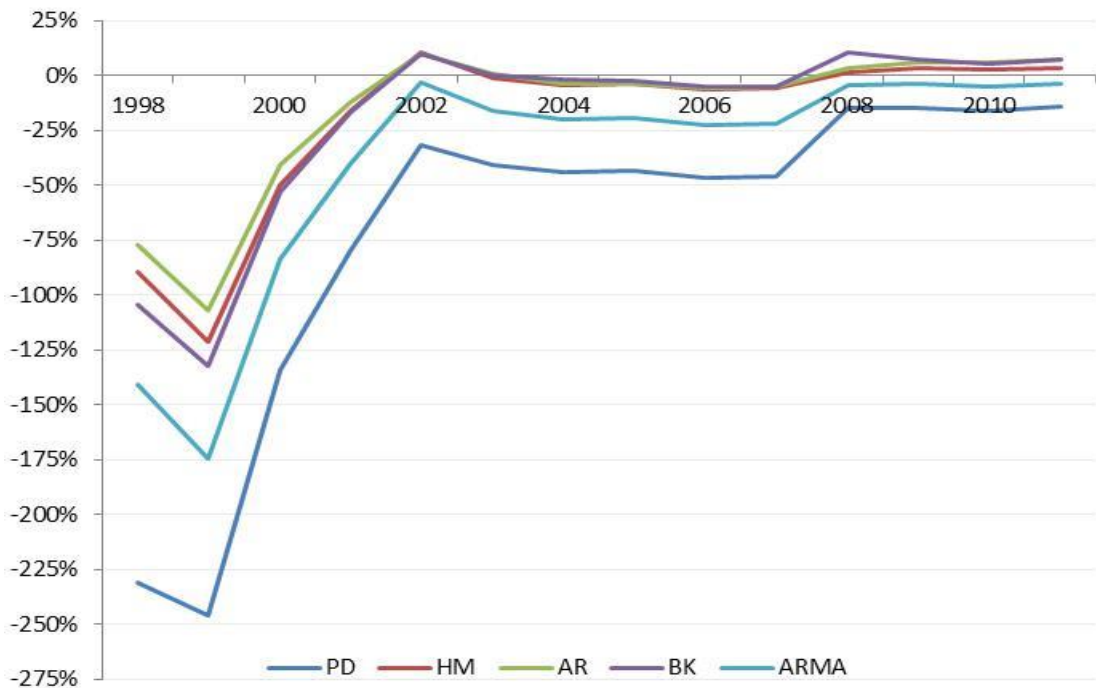
If we start accumulation of errors in 1963, we still observe stability in performance of the benchmark model and to less extent historical mean and AR adjustments. In turn, ARMA adjustment and regular price-dividend ratio deliver highly volatile  $R^2$ , though now they perform even better than other predictors for considerable period of time (1975-1994), but this performance is further destroyed in late 90's.

Figure 7. Evolution of OOS  $R^2$  for different predictors (accumulation starts in 1978)



*PD, HM, AR, BK and ARMA denote regular price-dividend ratio, long-term historical mean adjustment, AR adjustment, benchmark (Binsbergen and Koijen) model, and ARMA adjustment accordingly.*

Figure 8. Evolution of OOS  $R^2$  for different predictors (accumulation starts in 1993)



*PD, HM, AR, BK and ARMA denote regular price-dividend ratio, long-term historical mean adjustment, AR adjustment, benchmark (Binsbergen and Koijen) model, and ARMA adjustment accordingly.*

In the last two periods we can explicitly observe the extent of negative performance in late 90's for different predictors and their recovery speed. Regular price-dividend ratio generates enormous errors in this period (1995-1999), while adjusted ratios and the benchmark model compensate their errors partially by accounting for dividend growth expectations. In both periods it is only ARMA adjustment and regular price-dividend ratio that do not recover to the positive values by 2011.

Considering our out-of-sample analysis, we can make three most important conclusions. First of all, both regular price-dividend ratio and ARMA adjustment are not able to achieve robust out-of sample results and do not pass an out-of-sample check. They have periods when they generate smaller forecast errors than other predictors do, but this is not a systematic pattern. In long-term they fall below other predictors regardless of the period and sometimes they generate even larger errors than the historical mean of returns do. Second, historical mean and AR adjustments show a high degree of stability in their out-of-sample performance and if we exclude late 90's, they would deliver out-of-sample  $R^2$  around 13-14%, which is significant. This period (late 90's) was extremely unfavorable for all predictors due to rapidly growing price without substantial changes in dividends. As a consequence price-dividend ratio substantially increased, but forecasts of future dividends stayed at the same level. Thus, all models predicted negative returns, while in reality they were positive for several years generating large errors for all predictors and pushing out-of-sample  $R^2$  down. Finally, the benchmark model appears to be the best out-of-sample. The key advantage of the benchmark model is even faster stabilization in early years, though its performance is also negatively affected in late 90's.

### *6.3 Economic significance*

We construct a trading strategy as described in *Section 5.3* for all models and for four different initiation times: 1948, 1963, 1978 and 1993. All strategies are terminated in the end of sample period (2011). Exhibit 7 summarizes ex-post Sharpe ratios for both unconstrained strategy (i.e. with possible negative weight in the market portfolio) and constrained strategy, when the negative weight in the market is forbidden (i.e. it is set to zero, when the model predicts market return below the risk-free rate).

*Exhibit 7. Ex-post Sharpe ratio for trading strategies based on different models*

Unconstrained (allow for negative weight in market)							
Initiation	Market	Returns Historical Mean	PD	PD_HM	PD_AR	PD_ARMA	BK
1948	0.337	0.302	0.286	0.452	0.479	0.384	0.528
1963	0.235	0.199	0.177	0.366	0.395	0.306	0.425
1978	0.318	0.196	0.048	0.341	0.379	0.241	0.396
1993	0.017	0.166	-0.2	0.134	0.21	0.01	0.206
Constrained (negative weight in market is forbidden)							
Initiation	Market	Returns Historical Mean	PD	PD_HM	PD_AR	PD_ARMA	BK
1948	0.337	0.297	0.42	0.449	0.486	0.41	0.534
1963	0.235	0.192	0.334	0.36	0.403	0.34	0.43
1978	0.318	0.188	0.299	0.344	0.393	0.32	0.401
1993	0.017	0.166	0.087	0.13	0.227	0.094	0.222

*PD, PD\_HM, PD\_AR, BK and PD\_ARMA denote regular price-dividend ratio, long-term historical mean adjustment, AR adjustment, benchmark (Binsbergen and Koijen) model, and ARMA adjustment accordingly.*

As we expected, the results generally fall in line with out-of-sample estimation. AR adjustment and the benchmark model deliver the highest Sharpe ratios in all periods. For the longest period (1948-2011) their Sharpe ratios in case of unconstrained strategy are 0.479 and 0.528 respectively, which is significantly above Sharpe ratio for the market, historical mean returns forecast and regular price-dividend ratio – 0.337, 0.302 and 0.286 respectively. They appear to be especially beneficial in 1993-2011 period obtaining Sharpe ratios above 0.2, while the market comes up with 0.02. Constrained strategy allows achieving higher Sharpe ratio in most cases, which corresponds to common sense intuition. ARMA adjustment is not able to capture all benefits of the benchmark model and always deliver Sharpe ratios below historical mean adjustment and comparable with regular price-dividend ratio. *Exhibit 8* presents Certainty Equivalent returns (CE) for all strategies.

*Exhibit 8. Certainty Equivalent returns for trading strategies based on different models (%)*

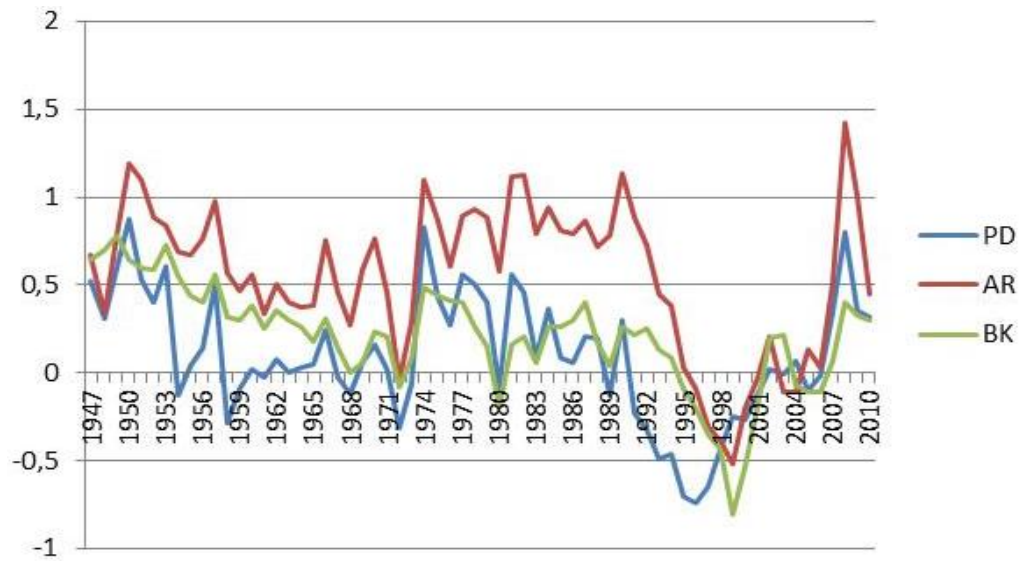
Unconstrained (allow for negative weight in market)							
Initiation	Market	Returns Historical Mean	PD	PD_HM	PD_AR	PD_ARMA	BK
1948	5.96	5.97	5.61	7.57	7.99	6.71	8.12
1963	4.79	5.57	5.49	7.25	7.64	6.55	7.75
1978	6.23	5.36	4.58	6.96	7.47	5.96	7.49
1993	1.66	2.00	0.37	2.77	3.61	1.80	3.52
Constrained (negative weight in market is forbidden)							
Initiation	Market	Returns Historical Mean	PD	PD_HM	PD_AR	PD_ARMA	BK
1948	5.96	5.94	6.38	7.51	8.05	6.90	8.17
1963	4.79	5.53	6.39	7.17	7.72	6.79	7.81
1978	6.23	5.29	6.19	6.98	7.63	6.61	7.64
1993	1.66	2.00	3.08	2.83	3.88	2.88	3.55

*PD, PD\_HM, PD\_AR, BK and PD\_ARMA denote regular price-dividend ratio, long-term historical mean adjustment, AR adjustment, benchmark (Binsbergen and Koijen) model, and ARMA adjustment accordingly.*

The highest CE is again achieved by AR adjustment and the benchmark model in all periods. For the unconstrained strategy in 1948-2011 period, they are 7.99% and 8.12% respectively. CE of the strategy based on regular price-dividend ratio equals to 5.61% in the same period. Intuitively, it means that an investor who tries to forecast returns with regular price-dividend ratio would be willing to pay up to 2.38% and 2.51% of his wealth to get an access to AR adjustment and the model of Koijen and van Binsbergen respectively. If we consider 1993-2011 period his willingness to pay will increase to 3.24% and 3.15%, which is obviously a substantial part of his wealth. It is still worth mentioning that trading strategy based on regular price-dividend ratio performs significantly better if negative weight in the market is forbidden, but it is still not able achieve sound performance of AR adjustment and the benchmark model.

As a final step of our analysis, we examine how market portfolio weights of the selected trading strategies evolve over time. Figure 9 presents market weights for unconstrained trading strategies based on AR adjustment, benchmark model and regular price-dividend ratio.

Figure 9. Evolution of market weights for different unconstrained trading strategies for the period 1948-2011



*PD, AR, and BK denote regular price-dividend ratio, AR adjustment, and benchmark (Binsbergen and Koijen) model accordingly.*

Considering the fact that all strategies are designed for the same investor (i.e. they all correspond to the same level of risk-aversion), we can claim that AR adjustment typically predicts the largest ex-ante market Sharpe ratio since it always suggests fairly aggressive market weights. In turn, regular price-dividend ratio comes up with very conservative strategy, which turns out to be unreasonable according to presented below ex-post Sharpe ratios and CE. Finally, it is the benchmark model, which delivers the highest SR and CE in most of the periods and its market weight typically lies between the weights proposed by AR adjustment and regular price-dividend ratio.

Overall we conclude that both the benchmark model and AR adjustment generate feasible trading strategies and lead to significant value gains regardless of the time span.



## 7. Conclusion

In our master's thesis we studied if forecasting of future dividends can improve the predictability of stock returns with price-dividend ratio. We analyzed this question from different perspectives including in-sample and out-of-sample analysis, as well as economic significance of results. Another dimension of our analysis was dedicated to comparison of two competing methodologies: latent variables approach in the spirit of Binsbergen and Koijen (2010) and adjusted price-dividend ratio proposed by Lacerda and Santa-Clara (2010). Since the question of stock returns predictability is very general and complex, we did not expect to get an exact and simple answer to our research questions. Nevertheless, we would like to focus on the key findings, which were discovered during our analysis.

First of all, we have shown that the theory behind adjusted price-dividend ratio methodology finds a convenient empirical support. By forecasting dividend growth with either historical mean or AR(1)-process, we were able to filter out some variation in price-dividend ratio due to variation in expected dividends. Thus, we obtained a substantial improvement in predictability of stock returns relative to regular price-dividend ratio. Moreover, we documented that AR adjustment always delivered slightly better results than historical mean adjustment proposed by Lacerda and Santa-Clara(2010). This difference in results became especially important when we had constructed simple trading strategies, where AR adjustment delivered more substantial benefits to investors. Moreover, out-of-sample results were robust and relative performance of different predictors was not sensitive to the choice of sample period.

The comparison latent-variables approach with adjusted price-dividend ratio methodology was two-fold. On the one hand, the model of Koijen and van Binsbergen did not show any improvements in predictability of stock returns above price-dividend ratio, while all adjusted ratios explained larger part of variation in stock market returns. On the other hand, the benchmark model was definitely the best from out-of-sample perspective, though AR adjustment was consistently close to it. Therefore, we conclude that there is a strong predictive potential behind the latent variables approach, which can be exploited in real-time delivering significant value gains for investors.

Quite surprisingly, ARMA adjustment, which is structurally equivalent to the model of Koijen and van Binsbergen have shown relatively poor performance, especially out-of-sample. For some sample periods it was not even able to outperform regular price-dividend ratio. We conclude that its mediocre performance is mostly driven by extra noise generated by inclusion of moving-average term into the realized dividend growth rate model. Moreover, we believe that the assumption that the market uses ARMA(1,1) model to forecast dividend growth rates could be overcomplicated, which was supported by our empirical results.

Finally, our results show that the aggregate stock market returns are predictable to reasonable extent both in- and out-of-sample. This predictability is significant, especially when we analyze actual value gains for trading strategies based on our predictors. We believe that there is still huge uncovered potential behind both adjusted price-dividend ratio and latent variables approach methodologies. In particular, one can estimate future dividends from the current market conditions instead of historical data. For example, Golez (2012) extracts dividend forecasts from the prices of futures and options traded in the market, which also leads to the adjusted price-dividend ratio, which captures significant part of variation in realized returns. If one can propose better and more precise way to forecast dividends, he would be able to improve returns predictability even further.

## Appendix A

We first reformulate the model in the standard state-space form. We define an expanded state vector as:

$$X_t = \begin{bmatrix} \hat{g}_{t-1} \\ \varepsilon_t^D \\ \varepsilon_t^g \\ \varepsilon_t^\mu \end{bmatrix}$$

which satisfies:

$$X_{t+1} = FX_t + \Gamma \varepsilon_{t+1}^X$$

with

$$F = \begin{bmatrix} \gamma_1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\Gamma = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\varepsilon_{t+1}^X = \begin{bmatrix} \varepsilon_t^D \\ \varepsilon_t^g \\ \varepsilon_t^\mu \end{bmatrix}$$

which are assumed to be jointly normally distributed.

The measurement equation with an observable vector  $Y_t = (\Delta d_t, pd_t)$ , is:

$$Y_t = M_0 + M_1 Y_{t-1} + M_2 X_t$$

with

$$M_0 = \begin{bmatrix} \gamma_1 \\ (1 - \delta_1)C \end{bmatrix}$$

$$M_1 = \begin{bmatrix} 0 & 0 \\ 0 & \delta_1 \end{bmatrix}$$

$$M_2 = \begin{bmatrix} 1 & 1 & 0 & 0 \\ B_2(\gamma_1 - \delta_1) & 0 & B_2 & -B_1 \end{bmatrix}$$

The Kalman Filter is then constructed as follows:

$$X_{0|0} = E[X_0] = \mathbf{0}_{4 \times 1}$$

$$P_{0|0} = E[X_0 X_0']$$

$$X_{t|t-1} = FX_{t-1|t-1}$$

$$P_{t|t-1} = FP_{t-1|t-1}F' + \Gamma \Sigma \Gamma'$$

$$\eta_t = Y_t - M_0 - M_1 Y_{t-1} - M_2 X_{t|t-1}$$

$$\eta_t = M_2 P_{t|t-1} M_2'$$

$$K_t = P_{t|t-1} M_2' S_t^{-1}$$

$$X_{t|t} = X_{t|t-1} + K_t \eta_t$$

$$P_{t|t} = (I - K_t M_2) P_{t|t-1}$$

The likelihood function is further computed based on prediction errors and their covariance matrix:

$$LL = - \sum_{t=1}^T \log(\det(S_t)) - \sum_{t=1}^T \eta_t' S_t^{-1} \eta_t$$

Finally, the covariance matrix of the shocks is:

$$\Sigma \equiv \text{var} \left( \begin{bmatrix} \varepsilon_{t+1}^d \\ \varepsilon_{t+1}^g \\ \varepsilon_{t+1}^\mu \end{bmatrix} \right) = \begin{bmatrix} \sigma_D^2 & \sigma_{Dg} & \sigma_{D\mu} \\ \sigma_{Dg} & \sigma_g^2 & \sigma_{\mu g} \\ \sigma_{D\mu} & \sigma_{\mu g} & \sigma_\mu^2 \end{bmatrix}$$

We further maximize the likelihood function with simulated annealing procedure.

## Appendix B

Simulated annealing is a probabilistic algorithm designed to solve global optimization problem. A common problem of conventional algorithms based on numerical derivatives is inability to leave the local maximum point when they reach it. In other words, with conventional algorithms one has to make a very good initial guess of parameters to be sure that the algorithm will reach global rather than local maximum point. Simulated annealing overcomes this problem. At each step it generates a random point in the function domain. If the value of function in this point is greater than currently reached maximum, it moves to this point. However, if it is less than the current maximum, it still can move to this point with the probability that is decreasing with number of iterations. It allows looking for global maximum points beyond the neighborhood of current maximum point. More formally, assume that  $X_0$  and  $f_0$  and currently reached maximum point and value of the function at this point respectively. We also have  $T$ , which is called a temperature parameter. At each step we generate a new random point  $X$  with multivariate normal distribution. We further compute  $f$  – value of the function at point  $X$ . Then, we search process moves from  $X_0$  to  $X$  with the following probabilities:

$$X_0 = \begin{cases} 1, & \text{if } f > f_0, \\ e^{\frac{f-f_0}{T}}, & \text{if } f \leq f_0 \end{cases}$$

Hence, the algorithm moves to the point with lower function value with probability, which declines with difference between the old and new value and also declines with  $T$ . Initially  $T$  is very high, which makes the algorithm chaotically jumping between the local maximum areas, however it gradually declines after fixed number of iterations (200 in our case) according to the pre-specified cooling schedule:

$$T_{new} = 0.9 * T_{old}$$

Algorithm also tracks currently reached global maximum, so it returns to this point each time temperature is reduced. Thus, as temperature decreases, the algorithm focuses on the most promising areas. The algorithm is terminated when temperature becomes below some pre-specified level. The drawback of the algorithm is its slow speed of convergence, which is fully compensated by its benefits when computational efficiency is not a major issue.

## Appendix C

This appendix contains all programming codes in MATLAB that were used for estimation of Binsbergen and Koijen model.

### Likelihood function:

```
function [LL] = Likelihood(parameters,ro_gD, Y, F, G, ro, cappa)
```

```
%Model formulation
```

```
F(1,1)=parameters(3);
```

```
A = cappa/(1-ro)+(parameters(1)-parameters(2))/(1-ro);
```

```
B1= 1/(1-ro*parameters(4));
```

```
B2= 1/(1-ro*parameters(3));
```

```
M0=[parameters(1); (1-parameters(4))*A];
```

```
M1=[zeros(1,2); 0 parameters(4)];
```

```
M2=[1 1 0 0; B2*(parameters(3)-parameters(4)) 0 B2 -B1];
```

```
COV=[parameters(7)^2 parameters(5)*parameters(7)*ro_gD
```

```
parameters(6)*parameters(7)*parameters(9);
```

```
parameters(5)*parameters(7)*ro_gD parameters(5)^2
```

```
parameters(6)*parameters(5)*parameters(8);
```

```
parameters(6)*parameters(7)*parameters(9)
```

```
parameters(6)*parameters(5)*parameters(8) parameters(6)^2];
```

```
% Computation
```

```
X_post=zeros(4,1);
```

```
P_post=X_post*X_post';
```

```
LL_pre=0;
```

```
for t=1:(length(Y(1,:))-1)
```

```
    X_pre=F*X_post;
```

```
    P_pre=F*P_post*F'+G*COV*G';
```

```
    Eta=Y(:,t+1)-M0-M1*Y(:,t)-M2*X_pre;
```

```
    S=M2*P_pre*M2';
```

```
    K=P_pre*M2'*inv(S);
```

```
    X_post=X_pre+K*Eta;
```

```
    P_post=(eye(4,4)-K*M2)*P_pre;
```

```
    LL_pre=LL_pre-log(det(S))-Eta'*inv(S)*Eta;
```

```
end;
```

```
LL=LL_pre;
```

```
end
```

### Estimation with simulated annealing:

```
X=[0.05 0.05 0.5 0.5 0.3 0.3 0.3 0 0];
```

```
V0=[.15 .15 .5 .5 .1 .1 .2 .1 1];
```

```
%Bounds for parameters are introduced to increase estimation speed
```

```
Bounds=[.001 .151; .001 .151; .001 .999; .001 .999; .001 .601; .001 .601; .001
```

```
.601; -.999 .999; -.999 -.999]
```

```
%Scaling vector for generation of new points
```

```
C=[1 1 7 7 4 4 4 13 13]
```

```

C=0.00176776.*C
eps=.00001;
N_S=200
T0=200;
rT=0.9;
F=zeros(4,4);
F(1,3)=1;
G = [ zeros(1,3); eye(3,3)];
ro = exp(mean(pd))/(1+exp(mean(pd)));
cappa = log (1+ exp(mean(pd)))-ro*mean(pd);
Y=[0 d'; pd'];
f=Likelihood(X, 0, Y, F, G, ro, cappa);
f_opt=f;
X_opt=X;
X_new=X;
T=T0;
V=V0;
D=zeros(1,9);
%Simmulated annealing
while T>eps
    D=T^(.5).*C;
    for i=1:1:N_S
        for k=1:1:4
            X_new(k)=normrnd(X(k), D(k));
            if (X_new(k)>Bounds(k,2))
                X_new(k)=Bounds(k,2);
            end
            if (X_new(k)<Bounds(k,1))
                X_new(k)=Bounds(k,1);
            end
        end
        f_new=Likelihood(X_new, 0, Y, F, G, ro, cappa);
        if (f_new>f)
            X=X_new;
            f=f_new;
        else
            if (exp((f_new-f)/T)>rand)
                X=X_new;
                f=f_new;
            end
        end
        if (f_new>f_opt)
            X=X_new;
            f=f_new;
            X_opt=X_new;
            f_opt=f_new;
        end
    end
end

```

```
X=X_opt;  
f=f_opt;  
T=T*rT;  
end
```



---

## Bibliography

- Alizamir, S., Rebennack, S., Pardalos, P., Improving the Neighborhood Selection Strategy in Simulated Annealing Using the Optimal Stopping Problem, Simulated Annealing, Cher Ming Tan (Ed.), ISBN: 978-953-7619-07-7, InTech, DOI: 10.5772/5571. Available from: [http://www.intechopen.com/books/simulated\\_annealing/improving\\_the\\_neighborhood\\_selection\\_strategy\\_in\\_simulated\\_annealing\\_using\\_the\\_optimal\\_stopping\\_problem](http://www.intechopen.com/books/simulated_annealing/improving_the_neighborhood_selection_strategy_in_simulated_annealing_using_the_optimal_stopping_problem)
- Ang, A., and J. Liu, 2004. How to discount cashflows with time-varying expected returns. *Journal of Finance*, 59: 2745–2783.
- Banzal, R., Yaron, A., 2004. Risks for the long-run: a potential resolution of asset pricing puzzles. *Journal of Finance*, 59(4), 1481-1509.
- Ball, Ray, 1978. Anomalies in relationships between securities' yields and yield-surrogates. *Journal of Financial Economics*, 6: 103-126.
- Bekaert, G., E. Engstrom, and S. R. Grenadier, 2001. Stock and bond pricing in an affine economy. NBER Working Paper No. 7346.
- Brennan, M. J., and Y. Xia, 2005. Persistence, predictability, and portfolio planning. Working Paper, UCLA.
- Burnside, C., 1998. Solving asset pricing models with gaussian shocks. *Journal of Economic Dynamics and Control*, 22: 329–340.
- Campbell, John Y., 1991. A variance decomposition for stock returns. *Economic Journal*, 101, 157-179.
- Campbell, John Y. and Robert Shiller, 1988. The dividend-price ratio and expectations of future dividends and discount factors. *Review of Financial Studies*, 1:195–227.
- Campbell, John Y. and Thompson, S., 2007. Predicting excess stock returns out of sample: can anything beat the historical average? *Review of Financial Studies* 21(4): 1509-1531.
- Cochrane, J. H., 1991. Explaining the variance of price-dividend ratios. *Review of Financial Studies*, 5(2): 243-280.
- Cochrane, J. H., 2007. The dog that did not bark: A defense of return predictability. *Review of Financial Studies*, 21 (4): 1533–1575.
- Cochrane, J. H., 2008. State-space vs. VAR models for stock returns. Working paper. University of Chicago.
- Dow, Charles H., 1920. Scientific stock speculation. *The Magazine of Wall Street* (New York).
- Fama, E. and French, 1988. Dividend yields and expected stock returns. *Journal of Financial Economics*, 22:3–27.

- Fama, E. and French, K., 2000. Disappearing dividends: changing firm characteristics or lower propensity to pay. *Journal of Financial Economics*, 60: 3-43.
- Fernandez-Villaverde, J., Rubio-Ramirez, J., and Santos, M., 2006. Convergence properties of the likelihood of computed dynamic models. *Econometrics*, 74: 93-119.
- Ferson, W. E., Sarkassian, S., Simin, T. T., 2003. Spurious regressions in financial economics? *Journal of Finance*, 58 (4), 1393-1413
- Flood, Robert P., Robert J. Hodrick, and Paul Kaplan, 1986. An evaluation of recent evidence on stock market bubbles. Unpublished manuscript (National Bureau of Economic Research, Cambridge, MA).
- Gabaix, X., 2007. Linearity-generating processes: a modeling tool yielding closed forms for asset prices. Working paper MIT.
- Goetzmann W. N., and P. Jorion, 1995. A longer look at dividend yields. *Journal of Business*, 68: 483-508.
- Goffe, W., Ferrier, G., Rogers, J., 1992. Global optimization of statistical functions with simulated annealing. *Journal of Econometrics*, 60: 65-99.
- Golez, B., 2012. Expected returns and dividend growth rates implied in derivative markets. Available at <http://ssrn.com/abstract=1271607>
- Gordon, Myron J., 1959. Dividends, earnings and stock prices. *Review of Economics and Statistics* (The MIT Press), 41 (2): 99-105.
- Goyal, A., and I. Welch, 2008. A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, 21: 1455-1508.
- Hamilton, J. D., 1994. *Time series analysis*. Princeton University Press, Princeton, NY.
- Hansen, L., Heaton, J., Li, N., 2008. Consumption strikes back? Measuring long-run risk. *Journal of Political Economy*, 116 (2): 260-302
- Harvey, A., 1989. *Forecasting, structural time series models and the Kalman Filter*. Cambridge University Press, Cambridge, UK.
- Julier, S., Uhlmann, J., 1997. A new extension of the Kalman Filter to nonlinear systems. *Proceedings SPIE Signal Processing, Sensor Fusion, and Target Recognition VI*, 3068: 182-193.
- Koijen R., and J. van Binsbergen, 2010. Predictive regressions: A present-value approach. *Journal of Finance*, 65 (4).
- Lacerda, F., and P. Santa-Clara, 2010. Forecasting dividend growth to better predict returns. Working Paper. Available at <http://docentes.fe.unl.pt/~psc>
- Lettau, M., and Ludvigson, S., 2005. Expected returns and expected dividend growth. *Journal of Financial Economics*, 76: 585-626.

- Lettau, M., and S. van Nieuwerburgh, 2008. Reconciling the return predictability evidence. *Review of Financial Studies*, 21 (4):1607–1652.
- Lewellen, J. W., 2004. Predicting returns with financial ratios. *Journal of Financial Economics*, 74: 209-235.
- Menzly, L., T. Santos, and P. Veronesi, 2004. Understanding predictability. *Journal of Political Economy*, 112(1):1–47.
- Pastor, L., and P. Veronesi, 2003. Stock valuation and learning about profitability. *Journal of Finance*, 58: 1749–1789.
- Pastor, L., and P. Veronesi, 2006. Was There a Nasdaq Bubble in the Late 1990s? *Journal of Financial Economics*, 81: 61–100.
- Pastor, L., and R. F. Stambaugh, 2006. Predictive systems: living with imperfect predictors. Working Paper, University of Chicago and CRSP.
- Rozeff, Michael, 1984. Stock prices and social dynamics. *Brookings Papers on Economic Activity*, 2: 457-498.
- Rytchkov, O., 2007. Filtering out expected dividends and expected returns. Unpublished paper, Texas A&M.
- Stambaugh, R., 1999. Predictive regressions. *Journal of Financial Economics*, 54: 375-421.
- Trojani, F., and Piatti I., 2012. Dividend growth predictability and the price-dividend ratio. Swiss finance institute research paper, 12-42. Available at SSRN: <http://ssrn.com/abstract=2077381>
- Yun, J., 2011. A present-value approach to variable selection. Unpublished Manuscript.

**Artem Abramov**  
**Polina Panfilova**

BI Norwegian Business School – Preliminary Thesis Report

# Adjusted price-dividend ratio and stock returns predictability

Thesis supervisor: Bruno Gerard

Hand-in date:  
15.01.2013

Campus:  
BI Oslo

Programme:

Master of Science in Financial Economics

Master of Science in Business and Economics with major in Finance

## Content

<b>1. INTRODUCTION.....</b>	<b>1</b>
<b>2. MODEL .....</b>	<b>6</b>
2.1 HISTORICAL MEAN ADJUSTMENT.....	7
2.2 ARMA ADJUSTMENT .....	8
2.3 AR ADJUSTMENT.....	9
<b>3. STATISTICAL METHODOLOGY .....</b>	<b>9</b>
<b>4. DATA .....</b>	<b>10</b>
<b>5. PRELIMINARY ESTIMATION RESULTS.....</b>	<b>11</b>
<b>6. BIBLIOGRAPHY .....</b>	<b>13</b>

## 1. Introduction

The question of predictability of stock returns has always played an important role in financial economics. Researchers have been concerned with major economic forces, which drive a capital gain process and implications of various macroeconomic shocks for the equity returns. Moreover, this question was the one of the great importance for ordinary participants of financial markets, because any evidence of predictability would generate feasible trading strategies and lead to better understanding of portfolio management. Deep study in this field has been started quite recently, because in 60-70s the efficient market hypothesis was assumed to reflect the reality by majority of researchers, so predictability of stock returns was considered to be impossible. Nevertheless, the hypothesis that dividend price ratio (D/P) forecasts returns has always existed among researchers and practitioners (Dow, 1920 and Ball, 1978). The intuition of the hypothesis is that dividends is high relative to stock prices when expected returns are high given that future dividend growth rates remain constant. This intuition comes directly from the famous Dividend Discount Model (DDM) initially proposed by Gordon in 1959. Number of researchers found statistical evidence (mostly for annual data) that support the hypothesis. See, for example, Rozeff (1984), Flood, Hodrick and Kaplan (1986) and Campbell and Shiller (1987). However, the paper of Fama and French (1988) is often considered to be fundamental in this field. They not only confirmed statistical significance of dividend yields for prediction of future returns, but also discovered that forecasting power increases with the return horizon and provided strong economical intuition to support their findings. Another impact on the empirical research of stock returns predictability came from the paper of Campbell and Shiller (1988). They proposed log-linearized present-value model, which produces a simple relation between log price-dividend ratio and expectations of log dividend growth rates and log returns at the price of small approximation error. Thus, they stress the fact that variation of price-dividend ratio may reflect not only changes in expected returns, but also changes in expected future dividends and we should not expect parsimonious results by using regular price-dividend ratio solely to predict future returns. This fact was also pointed out by Menzly, Santos and Veronesi (2004) and Goetzmann and Jorion (1995). Moreover, many researchers claim that dividends are highly persistent implying at least some persistency in dividend yields. Hence, returns predictability may be mistakenly supported due to this persistency. See, for

example, Powell, Shi and Smith (2004). These two arguments inevitably result in a conclusion that stock returns and dividend growth predictability are best studied jointly.

There were many other valuation ratios, whose ability to forecast future returns was tested. However, what makes dividend yield special is strong intuitive interpretation and decent theoretical basis behind it (DDM). In turn, it is not as easy to construct convenient economical arguments, for example, for book-to-market ratio to support the hypothesis that it predicts future returns. Hence, it is not a big surprise that since the publication of Campbell and Shiller paper (1988), many other papers with different versions of present-value model appeared. See Koijen and van Binsbergen (2010), Lacerda and Santa-Clara (2010, working paper), Trojani and Piatti (2012), Cochrane (2007), Lettau and Van Nieuwerburgh (2008), Pasto and Veronesi (2003), Pastor and Veronesi (2006), Bekaert, Engstrom and Grenadier (2001), Burnside (1998), Ang and Liu (2004), Brennan and Xia (2005) and Rytchkov (2007). In cases of Pasto and Veronesi (2003), Pastor and Veronesi (2006), Bekaert, Engstrom and Grenadier (2001), Ang and Liu (2004) and Brennan and Xia (2005) price-dividend ratio is presented as an indefinite integral of exponentially-quadratic terms making empirical part of the work much more involved relative to other papers. They employ either generalized method of moments or two-step procedure to estimate their model. In turn, Koijen and van Binsbergen (2010), Trojani and Piatti (2012), Rytchkov (2007) and Cochrane (2007) combine the present-value model with the assumption that both expected returns and expected dividend growth are latent variables that follow an exogenously-specified time-series model. Then, they assume normality of the shocks to estimate the model with the maximum likelihood. Finally, they use filtering techniques to uncover expected returns and expected dividend growth rates. Rytchkov (2007) and Cochrane (2007) focus more on methodology construction, derivation of main properties of state-space models, applicability of Kalman filter and relaxation of different assumptions and the consequences for estimation techniques. Koijen and van Binsbergen (2010) concentrate on the empirical side of the latent-variables approach and aggregate the whole history of the price-dividend ratio and dividend growth rates to deliver predictors for future returns and dividend growth rates. Since our master's thesis is closely related to the paper of Koijen and van Binsbergen (2010), we proceed to a more detailed discussion of their work.

Koijen and van Binsbergen (2010) model expected returns and expected dividend growth rates as latent variables, which follow low-order autoregressive processes. Following Pastor and Stamabaugh (2006) and Cochrane (2008) they assume that expected returns follow AR(1)-process, however they treat expected dividend growth rates differently depending on the choice of reinvestment strategy. Since they try to avoid effects of seasonality in dividend payments, they consider an annual model, which requires taking into account how dividends received within a particular year are reinvested. Two extreme reinvestment strategies are studied in detail. First, they reinvest dividends in 30-day T-bill and call it cash-invested dividends. Second, they reinvest dividends in the aggregate stock market and refer to it as market-invested dividends. Market-invested dividends appear to be far more volatile than cash-invested dividends supporting the fact that the choice of reinvestment strategy is extremely important. Interestingly enough, they assume that cash-invested expected growth rates are an AR(1)-process. By means of analytical argument this assumption implies that market-invested expected growth rates also exhibit moving average component and follow ARMA(1,1)-process. After specification of processes for latent variables, they employ log-linearization of realized returns in order to connect expected and realized variables through measurement equations. Then, they use Kalman filter not only to estimate unknown parameters, but also to filter out most likely values of latent variables. Later they find what fraction of realized returns and dividend growth variation can be explained by expected values, and compare these numbers to values of regular linear regressions with price-dividend ratio as a predictive variable. They discover that their model is superior to ordinary linear regression for both cash and market-invested dividends. Additionally, they emphasize that it is extremely important to study predictability of stock returns and dividend growth rates jointly because there is a tight relationship between the predictive coefficients of returns and dividend growth rates and the persistence of the dividend yield.

As long as log-linearized present-value model relates price dividend ratio to the expected returns and expected dividend growth rates, there is an alternative methodology, which does not require complex estimation techniques. The idea comes from Lacerda and Santa-Clara (2010), who argue that one can adjust price-dividend ratio for variations in expected growth rates and use the adjusted ratio to forecast future returns. However, this approach requires us to assume how market



participants estimate future dividend growth rates. Lacerda and Santa-Clara (2010) use moving average of historical growth rates as an estimate for future dividend growth. However, they reasonably stress the fact that the existence of better predictors is an open question and if such predictors are found, then they will presumably lead to better estimates of expected returns. Then, they transform an initial dividend-price ratio as follows:

$$x_t = dp_t + \frac{\bar{g}_t}{1 - \bar{\rho}_t},$$

where  $dp_t$ - log dividend-price ratio;

$\bar{g}_t$ - historical averaged log dividend growth;

$$\bar{\rho}_t = \frac{1}{1 + \left(\frac{D}{P}\right)_t}.$$

Finally, they use this new adjusted ratio as a predictive variable. The intuition behind this adjustment is to distinguish between change in dividend-price ratio due to changes in expected dividend growth rate and due to changes in expected future returns. They find out that adjusted dividend yield explains more variation in future returns than ordinary dividend yield both in- and out-of-sample.

The fact that Lacerda and Santa-Clara reveal statistically significant predictive power even out-of-sample becomes even more important in the light of Goyal and Welch (2008), who strongly criticize any evidence of returns predictability. They argue that in the real world we cannot use information that is not available yet. Good performance of some valuation ratios and other predictors in-sample is not practically important. Even if the true model exists and it is known, the true coefficients are unknown and we have to adjust their estimates as new data become available. Hence, for practical purposes Goyal and Welch measure out of sample performance and find that all common predictors show very poor results. In turn, Lacerda and Santa-Clara reach a certain degree of predictability even out-of-sample suggesting that the question of applicability in practice is not closed by Goyal and Welch.

In this master's thesis we will investigate if estimation of future dividends and dividends growth in the spirit of Lacerda and Santa-Clara (2010) leads to improvement of predictability of future equity returns. We will primarily focus on the practical aspects; thus, evaluation of the out-of-sample performance is defined as the main scientific question of the thesis. Our methodology basically coincides

with that used by Lacerda and Santa-Clara (2010); however, as was mentioned, they leave an open question of existence of better predictors for future dividend growth. Therefore, we not only replicate their procedure for our dataset, but also study two additional predictors. The first predictor is based on the assumptions of van Binsbergen and Kojien (2010) and is chosen due to parsimonious fit of the van Binsbergen's and Kojien's model. Under their assumptions realized dividend growth rates are ARMA(1,1)-process if cash-invested dividends are considered. In other words, we will test if our approach leads to the same results as a complex latent-variables model of Kojien and van Binsbergen and if simple OLS technique sufficient to study stock returns predictability and dividend growth rates jointly. The second predictor is an intermediate case, where we model realized dividend growth rates as the first order autoregressive process.

We forecast dividend growth rate and employ this forecast to construct an adjusted price-dividend ratio. Then, we test if this adjustment contributes significantly to predictability of future returns. As long as Lacerda and Santa-Clara used another dataset, we will replicate their procedure for our data to compare the performance of different predictors. Moreover, we will completely replicate the procedure of van Binsbergen and Kojien to use in-sample performance of their model as a benchmark. We will also extend their methodology to evaluate the performance of their model out-of-sample. The model will be re-estimated at each time step using only data available by that moment. It will allow us to construct out-of-sample measure of performance and compare it with our approach.

Finally, we will study how our predictors perform in the presence of additional predictive variables (output gap, term premium, different valuation ratios etc.). We study if our results hold internationally by considering LSE and TSE in the case if we are able to find sufficient data.

The rest of the preliminary master thesis report is organized as follows. In Section 2 we present our version of log-linearized present-value model. Section 3 is devoted to the empirical transformation of the model and description of estimation techniques. Section 4 contains description of the data. Section 5 demonstrates preliminary estimation results for in-sample performance.

## 2. Model

In this section we present a classical log-linearized present-value model initially proposed by Campbell and Shiller (1988)<sup>1</sup>. We follow the standard procedure to derive our model. Initially we assume that latent expected returns follow an AR(1)-process, though it can be extended for higher orders of ARMA-family. This assumption of persistency is consistent with findings of Fama and French (1988), Campbell and Cochrane (1999) and Kojien and van Binsbergen (2010). Then, following methodology of Lacerda and Santa-Clara we assume that parameters generating the process for the expected returns are known to investors, but the entire process for dividend growth is unknown to them, so they have to forecast it from the historical data. Finally, we assume that investors price the market in accordance with their estimates of future dividend growth. These assumptions allow us to simplify the model of Kojien and van Binsbergen. In other words, our model can be estimated with OLS, but potentially the results can be less parsimonious. The first objective of our work is to compare these results.

We use the following notation for log return, log price-dividend ratio and log dividend growth rate:

$$r_{t+1} \equiv \log\left(\frac{P_{t+1} + D_{t+1}}{P_t}\right),$$

$$pd_t = \log\left(\frac{P_t}{D_t}\right),$$

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

For expected returns we assume the following structure:

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

where  $\mu_t = E_t[r_{t+1}]$  and  $\delta_0$  is an unconditional mean of expected returns (if  $|\delta_1| < 1$ ).

Using the properties of autoregressive processes we can show that:

$$E_t[\mu_{t+k}] = \delta_0 + \delta_1^k(\mu_t - \delta_0)$$

Then, we proceed to linearization of the model for log returns.

$$r_{t+1} = \log\left(1 + \frac{P_{t+1}}{D_{t+1}}\right) + \Delta d_{t+1} - pd_t$$

---

<sup>1</sup> Different versions of this model were also presented by Cochrane (1991), Cochrane (2007), Van Nieuwerburgh (2006), Kojien and van Binsbergen (2010) and others.

Using a Taylor expansion around the historical average of  $pd_t$  up to time  $t$  and ignoring all terms after the first order, we get:

$$\begin{aligned} r_{t+1} &= \log(1 + \exp(pd_{t+1})) + \Delta d_{t+1} - pd_t = \\ &= \log(1 + \exp(\overline{pd_t})) + \frac{\exp(\overline{pd_t})}{1 + \exp(\overline{pd_t})} (pd_{t+1} - \overline{pd_t}) + o(pd_{t+1} - \overline{pd_t}) \\ &\quad + \Delta d_{t+1} - pd_t \cong \kappa_t + \rho_t pd_{t+1} + \Delta d_{t+1} - pd_t, \end{aligned}$$

where  $\overline{pd_t} = \sum_{i=1}^t \frac{pd_i}{t}$ ,  $\kappa_t = \log(1 + \exp(\overline{pd_t})) - \rho_t \overline{pd_t}$  and  $\rho_t = \frac{\exp(\overline{pd_t})}{1 + \exp(\overline{pd_t})}$ .

If we assume that the latter result is an exact equality rather than approximation we can iterate it and get:

$$\begin{aligned} pd_t &= \frac{\kappa_t}{1 - \rho_t} + \sum_{i=1}^{\infty} \rho_t^{i-1} (E_t[\Delta d_{t+i}] - E_t[r_{t+i}]) \\ &= \frac{\kappa_t - \delta_0}{1 - \rho_t} - \frac{\mu_t - \delta_0}{1 - \rho_t \delta_1} + \sum_{i=1}^{\infty} \rho_t^{i-1} E_t[\Delta d_{t+i}] \end{aligned}$$

Further transformation depends on the assumption of how investors forecast future dividend growth rates. We focus on three different alternatives.

## 2.1 Historical mean adjustment

First, following Lacerda and Santa-Clara we assume that

$$E_t[\Delta d_{t+i}] = \overline{g_{t,s}} = \sum_{j=t-s+1}^t \frac{\Delta d_j}{s},$$

Lacerda and Santa-Clara consider  $s=10$  as a reasonable sample size, because it roughly equals to a full business cycle. In order to completely replicate their procedure, we stick to this value of  $s$ . Given the assumption above, we can compute the last sum in expression for  $pd_t$ :

$$pd_t = \frac{\kappa_t - \delta_0}{1 - \rho_t} - \frac{\mu_t - \delta_0}{1 - \rho_t \delta_1} + \frac{\overline{g_{t,s}}}{1 - \rho_t}.$$

If we slightly rearrange the terms, we can derive an expression for the expected returns:

$$\begin{aligned} \mu_t &= \delta_0 + (1 - \rho_t \delta_1) \left( \frac{\kappa_t - \delta_0}{1 - \rho_t} + \frac{\overline{g_{t,s}}}{1 - \rho_t} - pd_t \right) \\ &= \delta_0 + (1 - \rho_t \delta_1) \left( \frac{\kappa_t - \delta_0}{1 - \rho_t} - pd_t^{HM} \right), \end{aligned}$$

where adjusted price-dividend ratio using historical mean forecast for dividend growth rates is equal to  $pd_t - \frac{\overline{g_{t,s}}}{1-\rho_t}$ .

## 2.2 ARMA adjustment

Second option evolves exactly from the model of van Binsbergen and Kojen for cash-invested dividends. We believe that cash-invested dividends reflect the reality much better than market-invested dividends, because in the real world dividends are often used for consumption rather than for reinvestment, which is equivalent to reinvestment in risk-free T-bill.

In state-space setup expected dividend growth follows AR(1)-process, while realized dividend growth is decomposed to its mean and orthogonal shock. More formally,

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

$$\Delta d_{t+1} = g_t + \varepsilon_{t+1}^D,$$

where  $g_t = E_t[\Delta d_{t+1}]$ . This model (see Cochrane (2008)) is exactly equivalent to ARMA(1,1) model for realized dividend growth. That is,

$$\Delta d_{t+1} = \gamma_0 + \gamma_1(\Delta d_t - \gamma_0) + (\beta - \gamma_1)\eta_t + \eta_{t+1},$$

where  $\eta_t$ — new error term (not equal to  $\varepsilon_t^D$ ).

Though an expression for  $\beta$  can be derived analytically, it is quite complex and does not exhibit any intuitive interpretation; therefore, we leave it as an unknown parameter. It is completely sufficient for estimation purposes. If investors forecast future dividend growth according to this model, then these forecasts are as follows:

$$f_{t,i}^s = E_t[\Delta d_{t+i}] = \gamma_0 + \gamma_1^i(\Delta d_t - \gamma_0) + \gamma_1^{i-1}(\beta - \gamma_1)\eta_t,$$

where  $s$  stands for number of past observations used for estimation of the model. For example,  $s=20$  means that at each step ARMA(1,1) is estimated using last 20 observations. Additionally, we consider  $f_{t,i}^*$ , which employ all available observations at time  $t$ . The initial expression for  $pd_t$  can be simplified again:

$$pd_t = \frac{\kappa_t + \gamma_0 - \delta_0}{1 - \rho_t} - \frac{\mu_t - \delta_0}{1 - \rho_t \delta_1} + \frac{\gamma_1(\Delta d_t - \gamma_0) + (\beta - \gamma_1)\eta_t}{1 - \rho_t \gamma_1},$$

$$\mu_t = \delta_0 + (1 - \rho_t \delta_1) \left( \frac{\kappa_t - \delta_0}{1 - \rho_t} + \frac{\gamma_0}{1 - \rho_t} + \frac{\gamma_1(\Delta d_t - \gamma_0) + (\beta - \gamma_1)\eta_t}{1 - \rho_t \gamma_1} - pd_t \right)$$

$$= \delta_0 + (1 - \rho_t \delta_1) \left( \frac{\kappa_t - \delta_0}{1 - \rho_t} - pd_t^{ARMA} \right)$$

### 2.3 AR adjustment

Finally, we consider an AR(1)-process for realized dividend growth.

$$\Delta d_{t+1} = \gamma_0 + \gamma_1(\Delta d_t - \gamma_0) + \eta_{t+1},$$

In this setup,  $f_{t,i}^s = E_t[\Delta d_{t+i}] = \gamma_0 + \gamma_1^i(\Delta d_t - \gamma_0)$

$$pd_t = \frac{\kappa_t + \gamma_0 - \delta_0}{1 - \rho_t} - \frac{\mu_t - \delta_0}{1 - \rho_t \delta_1} + \frac{\gamma_1(\Delta d_t - \gamma_0)}{1 - \rho_t \gamma_1},$$

$$\begin{aligned} \mu_t &= \delta_0 + (1 - \rho_t \delta_1) \left( \frac{\kappa_t - \delta_0}{1 - \rho_t} + \frac{\gamma_0}{1 - \rho_t} + \frac{\gamma_1(\Delta d_t - \gamma_0)}{1 - \rho_t \gamma_1} - pd_t \right) \\ &= \delta_0 + (1 - \rho_t \delta_1) \left( \frac{\kappa_t - \delta_0}{1 - \rho_t} - pd_t^{AR} \right) \end{aligned}$$

### 3. Statistical methodology

The intuition behind our further methodology is quite straightforward: we construct adjusted price-dividend ratio using one of the aforementioned assumptions for dividend growth rates. We believe that such modification will allow us to distinguish changes in dividend price-ratio due to changes in dividend growth rates and due to changes in expected returns. Therefore, we consequently run the following predictive regressions:

$$\begin{aligned} r_{t+1} &= \alpha_0^{HM} + \alpha_1^{HM} pd_t^{HM} + \varepsilon_{t+1}^{HM}, \\ r_{t+1} &= \alpha_0^{ARMA} + \alpha_1^{ARMA} pd_t^{ARMA} + \varepsilon_{t+1}^{ARMA}, \\ r_{t+1} &= \alpha_0^{AR} + \alpha_1^{AR} pd_t^{AR} + \varepsilon_{t+1}^{AR}, \end{aligned}$$

and

$$r_{t+1} = \alpha_0 + \alpha_1 pd_t + \varepsilon_{t+1}$$

Then, we analyze the results by examination of t-statistics and  $R^2$  values. However, it should be noted that this procedure will reflect only in-sample performance of the predictive variables. Therefore, we will also study out-of-sample performance. We leave the latter for the final master's thesis, but we describe methodology now. At each step we use only available information up to time t to estimate regression coefficients and forecast the return for the next period. There are different techniques of comparison of out-of sample performance, but we stick to the most common one (following Goyal and Welch (2008)). Out-of-sample  $R^2$  indicates how well the predictor performs relative to the historical mean of observations up time t. It is defined as follows:

$$R^2 = 1 - \frac{MSE_R}{MSE_M},$$

where  $MSE_R$  is the mean squared error from predictive regressions  $MSE_M$  is the mean squared error from historical mean.

#### 4. Data

We obtain S&P 500 historical prices and dividend data from Robert Shiller's website. Our sample is from 1926 to 2012; however, we can go back as far as 1871 for yearly data. Since we would like to incorporate dividend reinvestment strategy within a particular year, we cannot afford the latter, because dividend data from 1871 to 1925 are obtained from yearly data with linear interpolation to monthly values. Hence, the consideration of reinvestment strategy for this period will not lead to a precise result. Moreover, dividends in Robert Shiller's data set are given as 12 months moving sums of dividends paid on S&P 500. Thus, we transform initial series to monthly dividends. We should also mention that even from 1926 monthly dividend data is interpolated from S&P 500 quarter total returns; hence, in order to construct yearly dividend growth series, we reinvest dividends received in a particular quarter into 3-month T-bills. The T-bill rates from 1934 to 2012 are from Federal Reserve homepage. In order to estimate T-bill rates from 1926 to 1933, we regress T-bill rate on Commercial Paper rate from 1934 to 1971. The latter was obtained from the National Bureau of Economic Research (NBER) Macroeconomic history database. The regression yields the following estimation results:

$$T - \text{bill rate} = -0.0041 + 0.9307 * CPR$$

with an  $R^2$  of 98.04%. Therefore, we use the estimation results to construct synthetic T-bill rates from 1926 to 1933.

Finally, we use the aforementioned data to construct annual series for  $r_t$ ,  $\Delta d_t$  and  $pd_t$  from 1927 to 2012.

	$r_t$	$\Delta d_t$	$pd_t$
Mean	9.22%	4.40%	3.33
Standard deviation	18.67%	11.38%	0.46

It is worth mentioning that dividend growth in our sample is substantially more volatile and has smaller mean than in the sample used by van Binsbergen and Koijen. If we restrict our sample to 1946-2007 period, discrepancy in descriptive statistics almost disappears, but still remain, because they use more detailed data to construct dividend growth series (monthly rather than quarterly)

---

and exhaustive value-weighted index of all NYSE, Amex and Nasdaq stocks available from CRSP.

In the final version of this master's thesis we will also use the aforementioned data from CRSP, while current estimation results will be used as a brief overview.

## 5. Preliminary estimation results

We proceed to evaluation of in-sample performance of different models for dividend growth. We consider all three suggested adjustments and vary  $s$  (number of years used to estimate the model for dividend growth). Also in each case we consider the model with all available data used to estimate future dividend growth (denoted by  $s=*$ ). All predictive regressions estimated with OLS in EViews. The results are summarized in Exhibit 1. It should be noted that t-statistics were computed with OLS standard errors without any adjustments. However, it was reasonable, because neither heteroscedasticity nor autocorrelation was detected in fitted residuals in all regressions at the conventional significance levels (1%, 5%, 10%). The results for AR(1) and ARMA(1,1) models of dividend growth are presented only for  $s=*$ , because when we use rolling window approach,  $R^2$  values and t-statistics are essentially zero (even when  $s=30$ ). Moreover, there is no evidence so far to use AR and ARMA models with time-varying coefficients and such poor performance confirms that we simply create noise when allow for time-varying.

At first glance, we can conclude three things. First of all, we see that  $pd_t^{ARMA}$  and  $pd_t^{AR}$  exhibit weak performance when we start the sample from 1937, but fit changes drastically if we shift starting point of predictive regressions to 1947. It is related to the fact that 10-20 observations are very low numbers to deliver stable estimates of AR and ARMA models. Therefore, between 1937 and 1947 both  $pd_t^{ARMA}$  and  $pd_t^{AR}$  show enormous variation, but then they stabilize. In turn,  $pd_t^{HM}$  stabilizes much faster and performs good even in the long subsample. Second, if we ignore the exception above, all three models clearly outperform simple predictive regression with regular log price-dividend ratio. Therefore, we confirm results of Lacerda and Santa-Clara (2010) and Kojien and van Binsbergen (2010). Finally, there is no model, which is obviously more parsimonious than others. Still it is worth mentioning that for  $s=*$  all three models almost replicate each other when we consider the short subsamples (from 1947).



Hence, we can conclude that there is definitely a potential for the suggested models; however, detailed out-of-sample performance evaluation is required to make further conclusions.

---

## 6. Bibliography

- Ang, A., and J. Liu, 2004. How to discount cashflows with time-varying expected returns. *Journal of Finance*, 59: 2745–2783.
- Ball, Ray, 1978. Anomalies in relationships between securities' yields and yield-surrogates. *Journal of Financial Economics*, 6: 103-126.
- Bekaert, G., E. Engstrom, and S. R. Grenadier, 2001. Stock and bond pricing in an affine economy. NBER Working Paper No. 7346.
- Brennan, M. J., and Y. Xia, 2005. Persistence, predictability, and portfolio planning. Working Paper, UCLA.
- Burnside, C., 1998. Solving asset pricing models with gaussian shocks. *Journal of Economic Dynamics and Control*, 22: 329–340.
- Campbell, John Y. and Robert Shiller, 1987. The dividend-price ratio and expectations of future dividends and discount factors. Unpublished manuscript (Princeton University, Princeton, NJ).
- Campbell, John Y. and Robert Shiller, 1988. The dividend-price ratio and expectations of future dividends and discount factors. *Review of Financial Studies*, 1:195–227.
- Cochrane, J. H., 2007. The dog that did not bark: A defense of return predictability. *Review of Financial Studies*, 21 (4):1533–1575.
- Cochrane, J. H., 2008. State-space vs. VAR models for stock returns. Working paper. University of Chicago.
- Dow, Charles H., 1920. Scientific stock speculation. *The Magazine of Wall Street* (New York).
- Fama, E. and Kenneth French, 1988. Dividend yields and expected stock returns. *Journal of Financial Economics*, 22:3–27.
- Flood, Robert P., Robert J. Hodrick, and Paul Kaplan, 1986. An evaluation of recent evidence on stock market bubbles. Unpublished manuscript (National Bureau of Economic Research, Cambridge, MA).
- Goetzmann W. N., and P. Jorion, 1995. A longer look at dividend yields. *Journal of Business*, 68: 483-508.
- Gordon, Myron J., (1959). Dividends, earnings and stock prices. *Review of Economics and Statistics* (The MIT Press), 41 (2): 99-105.
- Goyal, A., and I. Welch, 2008. A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, 21: 1455-1508.
- Koijen R., and J. van Binsbergen, 2010. Predictive regressions: A present-value approach. *Journal of Finance*, 65 (4).
- Lacerda, F., and P. Santa-Clara, 2010. Forecasting dividend growth to better predict returns. Working Paper. Available at <http://docentes.fe.unl.pt/~psc>
- Lettau, M., and S. van Nieuwerburgh, 2008. Reconciling the return predictability evidence. *Review of Financial Studies*, 21 (4):1607–1652.

Menzly, L., T. Santos, and P. Veronesi, 2004. Understanding predictability. *Journal of Political Economy*, 112(1):1–47.

Pastor, L., and P. Veronesi, 2003. Stock valuation and learning about profitability. *Journal of Finance*, 58: 1749–1789.

Pastor, L., and P. Veronesi, 2006. Was There a Nasdaq Bubble in the Late 1990s? *Journal of Financial Economics*, 81: 61–100.

Pastor, L., and R. F. Stambaugh, 2006. Predictive systems: living with imperfect predictors. Working Paper, University of Chicago and CRSP.

Rozeff, Michael, 1984. Stock prices and social dynamics. *Brookings Papers on Economic Activity*, 2: 457-498.

Rytchkov, O., 2007. Filtering out expected dividends and expected returns. Unpublished paper, Texas A&M.

Trojani, F., and Piatti I., 2012. Dividend growth predictability and the price-dividend ratio. Swiss finance institute research paper, 12-42. Available at SSRN: <http://ssrn.com/abstract=2077381>

*Exhibit 1. Evaluation of in-sample performance*

This table summarizes the results of estimation of different predictive regressions. Each regression includes next period log return as a dependent variable, while four different subsamples are considered for each predictor (1937-2007, 1937-2012, 1947-2007 and 1947-2012). The detailed description of different adjustments for price-dividend ratio is presented in Section 2. "s" denotes a number of previous observations used for estimation of next period log dividend growth. "s="\*" corresponds to the case when all available by the current moment observations are used for estimation of next period log dividend growth. For each predictive regression three values are presented:  $R^2$ , slope and t-statistic. "\*" inside the t-statistic's field reflects the significance at the 5%-level, while "\*\*\*" denotes the significance at the 1%-level. The best predictor in terms of either  $R^2$  or t-statistic is highlighted for each subsample (bold font).

Predictor	s	1937-2007			1937-2012		
		$R^2$	Slope	t-stat	$R^2$	Slope	t-stat
$pd_t^{HM}$	10	<b>14.29%</b>	-0.0785	<b>-3.3913**</b>	<b>11.94%</b>	-0.0775	<b>-3.1682**</b>
$pd_t^{HM}$	20	-	-	-	-	-	-
$pd_t^{HM}$	*	7.60%	-0.1102	-2.3821*	7.29%	-0.1169	-2.4121*
$pd_t^{ARMA}$	*	1.70%	-0.0408	-1.0929	1.77%	-0.0454	-1.1546
$pd_t^{AR}$	*	0.38%	0.0126	0.5118	0.38%	0.0139	0.535
$pd_t$		4.98%	-0.0768	-1.9009	6.73%	-0.0932	-2.3116*
Predictor	s	1947-2007			1947-2012		
		$R^2$	Slope	t-stat	$R^2$	Slope	t-stat
$pd_t^{HM}$	10	16.18%	-0.0871	-3.3751**	13.27%	-0.0865	-3.1290**
$pd_t^{HM}$	20	15.07%	-0.1283	-3.2355**	12.20%	-0.1237	-2.9819**
$pd_t^{HM}$	*	17.32%	-0.1721	-3.5159**	<b>15.78%</b>	-0.182	<b>-3.4624**</b>
$pd_t^{ARMA}$	*	16.88%	-0.1673	-3.4610**	15.26%	-0.1763	-3.3950**
$pd_t^{AR}$	*	<b>17.50%</b>	-0.161	<b>-3.5373**</b>	12.73%	-0.1514	-3.0556**
$pd_t$		9.05%	-0.0986	-2.4235*	10.90%	-0.1165	-2.7981**