



BI Norwegian Business School - campus Oslo

GRA 19703

Master Thesis

Thesis Master of Science

Predictability of Stock Returns: An application of present-value state-space models to the German Stock Market

Navn: Yannik Lux

Start: 15.01.2019 09.00

Finish: 01.07.2019 12.00

**Predictability of Stock Returns:
An application of present-value state-space models to
the German Stock Market**

Master Thesis

Hand-in date:

July 1, 2019

Supervisor: Patrick Konermann

BI Norwegian Business School

Oslo, Norway

Author:

Yannik Lux

Master of Science in Business - QTEM Masters Network

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.

Abstract

In my thesis, I introduce a state-space representation of the present-value model to analyze predictability in the aggregated German stock market. The proposed model uses the information contained in annualized price-dividend ratios and realized dividend growth rates and defines relations to the latent state variables in the form of expected returns and expected dividend growth rates. I apply the Kalman Filter to generate estimates of the model parameters using a conditional Maximum Likelihood Estimation. The corresponding optimization problem is solved via an adjusted version of the Simulated Annealing algorithm. The final model produces good estimates for dividend-growth rates, while it lacks quality in terms of the estimation of stock returns.

Contents

1	Introduction	1
2	Literature Review	3
3	Theory & Research Approach	9
4	Data	10
5	Methodology	15
5.1	The Present-Value Identity	15
5.2	The Present-Value Model	19
5.3	State Space Models for Time Series Analysis	22
5.3.1	The Present-Value State-Space Model	23
5.4	The Kalman Filter	25
5.4.1	Maximum Likelihood Estimation of the Model Parameters	28
5.5	Estimation Restrictions	29
5.6	Simulated Annealing	30
6	Results	35
6.1	Estimation Results	35
6.2	In-Sample Performance	36
6.3	Hypothesis Testing	38
6.4	Validity of the Results	41
7	Conclusion	43

Bibliography	45
Appendices	50
A The Derivation of the Present-Value Model	50
B The Kalman Filter	54
B.1 Forecasting Y_t	56
B.2 Updating the State	57
B.3 Forecasting the State	59
B.4 Initialization Values	60
B.5 Summary of the Kalman Filter for the Present-Value Model	63
B.6 Maximum Likelihood Estimation of the Model Parameters	65
C The Kronecker Product	68
D Estimation Results for the CRSP Data Set	69
D.1 Results for the CRSP time series from 1946–2018	69
D.2 Replication: Comparison of the Results with the Present-Value model by Van Binsbergen and Koijen (2010)	70
E Likelihood-Ratio Test Results	71

1 Introduction

The existence of stock return predictability has been of major interest over decades. If a certain grade of predictability in the return of stocks would be present, it could be used to generate portfolio strategies that could gain abnormal returns for the investors. Furthermore, the research in stock predictability goes hand in hand with the search of factors that affect stock prices in the first place. Therefore, the studies of stock prediction also build up and enhance the understanding of important relations and drivers of stock returns.

Various approaches have been introduced to identify predictability. The results differ substantially from paper to paper. While some publications argue that there is no predictability at all, others find significant evidence for it. Fundamentals such as dividends appear to explain a big part of the variation of stock returns. More specifically, there seems to be a relation between the price-dividend ratio, expected returns and expected dividend growth. This relation was investigated in the present-value model of Campbell and Shiller (1988b) and fostered the research in stock and dividend predictability.

The state-space model of Van Binsbergen and Koijen (2010) builds up on Campbell's present-value identity and combines it with latent variables that follow simple time-series processes. In their concept, Binsbergen and Koijen make use of the Kalman filter and obtain good estimates via a maximum likelihood estimation. They find time variation and persistence in the expected returns and dividend growth rates, which contradicts the popular assumptions of constant expected returns. They achieve remarkable results considering the quality of the fit and of the out-of-sample predictions in the US stock market.

In my thesis, I replicate and derive the state-space representation of the present-value model according to Van Binsbergen and Koijen (2010) and apply it on the aggregated German Stock market. I introduce the Kalman Filter and subsequently

estimate the model parameters via an adjusted version of the Simulated Annealing algorithm. This is done in order to see if the patterns of time variation and persistence in expected stock returns can also be found in the German stock market. I compare the results to the reported findings on the aggregated American stock market. Ultimately, I compare the model's capability to estimate stock returns and dividend growth rates in comparison to other common methods.

2 Literature Review

Stock predictability has been an intensively researched and controversial topic for a long time. During the last century, many famous researchers discussed multiple methods, and many doubted the sheer existence of stock predictability. A classic view assumes stock returns to be close to be unpredictable. Expected returns and stock market volatility are not supposed to vary much over time (Cochrane, 2009a). This classical view is closely linked to guiding principles like the random-walk theory, the capital asset pricing model (CAPM) and the efficient market hypothesis (Fama, 1965), which suggest that stock prices reflect all available information and follow an unforeseeable path. Consequently, expected stocks returns were often assumed to be constant when formulating asset pricing models. Ultimately, these theories were initially seen as incompatible with the presence of return predictability.

However, many of these initial beliefs were scrutinized through new empirical research. Researchers found traces of predictability in stock returns, at least in the long-term, and the volatility was considered as changing over time. A broader spectrum of literature reports evidence of predictability in stock returns while not necessarily contradicting the classic financial theories. Often, the researchers instead try to combine them with their own views and empirical results. Fama (1991) claims in his review of previous work on market efficiency and predictability that expected returns are time-varying, persistent, and show signs of predictability. However, Fama also states that his findings are no conclusive evidence against efficient markets. Markets remain to be reasonably competitive and therefore, also quite efficient to some extent.

Subsequent research produced several equilibrium models that assume market efficiency while allowing for time-variation in expected stock returns. These models capture for example the effect of varying risk-aversion (Campbell and Cochrane, 1999), aggregate consumption risk (Bansal and Yaron, 2004) or variation in beliefs

on expected returns (Timmermann, 1993). De Cesari and Huang-Meier (2015) analyze the impact of private information and find a clear relation between returns and dividend growth rates. They conclude that managers actively use the information on stock prices to steer their dividend payout policies. It should be noted, that asset return-predictability is not necessarily a sign of inefficient markets anymore, and nowadays, a significant part of the literature reports time variation in expected returns as given.

In the 1980s, more and more empirical publications revealed the predictability of stock returns via financial ratios. Measures like the price-earning ratio, long-term-short-term bond-yield-spreads, macroeconomic variables or corporate decision variables showed forecasting abilities and received much interest in the field. Bollerslev et al. (2014) tries to predict aggregated stock market returns via the variance risk premium, defined as the difference between the risk-neutral and statistical expectations of the future return variance. He finds significant evidence of predictability using this measure.

One of these financial ratios, the price-dividend ratio, became a popular research subject in the literature (Ball, 1978; Campbell and Shiller, 1988a; Lewellen, 2004). The increasing interest in the interdependence between the dividend-price ratio and expected returns led other scientists to investigate the relation between these variables further. The discussion followed the basic intuition behind the renowned Gordon Growth Model or Dividend-Discount Model (Gordon, 1959). The model assumes that asset prices are worth the sum of all their future discounted dividends. This implies that stock prices move according to changes in expected future cash flows. If a stock is undervalued or, in other words, its price is relatively low compared to future dividends, the price is expected to rise, generating higher returns subsequently. Multiple publications apply the price-dividend ratio and prove its usefulness. Several practitioners found significant evidence of return predictability in simple uni-variate dividend-price ratio regression models (Campbell and Shiller,

1988b; Cochrane, 1992; Fama and French, 1988). Asimakopoulos et al. (2017) also find a significant forecasting capability of the dividend-price ratio for future dividend growth rates. While most research aggregates data on an annual base because of payout-policy and seasonality issues, they analyze monthly dividend data instead of annual observations and argue that time aggregation erases important information about the data. Stambaugh (1999) examines the power of predictive regression models in detail. In his paper, he also makes use of the dividend-price ratio to forecast future excess returns and further creates a trading strategy for investors based on it. Ang and Bekaert (2006) further investigate the predictive power of the dividend-yield and run several regression across multiple markets. While the dividend yield as a sole regressor shows no capability of predicting excess returns, they find that adding a second variable in the form of short-term interest rates results in a bi-variate regression with significant predictive power. In the article of Wachter and Warusawitharana (2015), the investors are even assumed to doubt predictability of returns, but change their mind when they get confronted with the predictive power of the price-dividend ratio.

Building upon the price-dividend ratio and the return-identity, John Y. Campbell and Robert J. Shiller introduced the so-called present-value model that attempts to capture the dynamic relations between the stock price movements, the dividend-price ratio, expected dividend growth rates and discount rates (1988a). An abstracted version as mentioned in Cochrane (2009a) can be presented as follows:

$$p_t - d_t = a + E \sum_{j=1}^{\infty} c^{j-1} (\Delta d_{t+j} - r_{t+j})$$

where $p_t - d_t$ is the log-price-dividend ration, Δd_t the log-dividend growth, r_t the log-return, and a and c are constant terms (a more detailed description is provided in the section 5). It implies that high prices must, mechanically, come from high future dividend growth or low future returns. Considering the decomposed variance of the price-dividend ratio, the ratio itself can only vary if either returns or dividend

growth are forecastable Cochrane (2009a). This approach, in particular, became the foundation of several publications in recent years. According to this identity, the analysis of general predictability in stock markets can be transformed into the question if dividend growth or returns are predictable. For example, Cochrane (2007, 2011) analyzes the movement of the price-dividend ratio and finds evidence for return-predictability, but not necessarily for dividend-growth predictability. Ang and Bekaert (2006) also employ a present-value model in their paper. They find that discount rates and short-term interest rates explain variation in the dividend-price ratio.

The literature suggests many different settings for the present-value model with varying assumptions, definitions and estimation methods. E.g. Pástor and Veronesi (2003; 2006) define the price-dividend ratio as an infinite sum or indefinite integral of quadratic terms. Bekaert and Grenadier (2001) and Ang and Liu (2004) estimate their model parameters via the generalized methods of moments. Lettau and Van Nieuwerburgh (2007) define a linearized present-value model and derive their parameters from reduced-form estimators. Thereby, they propose the critical assumption that expected growth rates and expected returns are equally persistent.

Jules van Binsbergen and Ralph Koijen's approach (2010) makes use of the price-dividend ratio and the present-value equation in the form of a state-space model. The state-space model comes along with the introduction of latent variables. Latent variables represent variables, which cannot be observed directly but can be derived via predefined relations to observed measurements. In their paper, these inferred variables are represented by expected returns and expected dividend growth rates, which are related to the price-dividend ratio, realized returns and realized dividend growth. The latent variables in a state space model can be estimated via the Kalman filter (Hamilton, 1994), which has been successfully applied in multiple return prediction models (see also Brandt and Kang (2004), Pástor and Stambaugh (2009), and Rytchkov (2012)). Koijen and Binsbergen further consider two different rein-

vestment strategies for the dividend payouts of the analyzed stocks that should have a considerable effect on their estimates and results. They consider dividends that are reinvested at the risk-free rate and dividends reinvested in the stock market. The impact of different reinvestment strategies in combination with the price-dividend ratio has previously been investigated by Chen (2009). They find time variation and persistence in the expected returns and dividend growth rates, which contradicts the popular assumptions of constant expected returns. They achieve remarkable results considering the quality of the fit and the out-of-sample predictions in the US stock market.

There are also other publications which work with state-space representations of the present-value identity. For example, Piatti and Trojani (2017) also introduces a state-space representation to model expected returns and dividends. In their approach, the model contains time-varying risk instead of the homoscedastic constant risk as it was assumed in Binsbergen & Koijen's paper. As a result, they find different outcomes in terms of the persistence in the latent variables, but also confirm evidence of predictability in the stock market in the end. In a follow-up paper, Piatti and Trojani (2019) develop an asymptotic testing method that further confirms these findings.

However, there is also a range of literature that doubt the sheer existence of predictability. Also more recent publications such as the ones of Goyal & Welch (2003, 2008) or Yongok Choi and Park (2016) criticize the capability of return forecasting models. Some of the research results of the corresponding models revealed some flaws in the measures. Dividend growth rates were commonly seen as hard to forecast, and the empirical findings of several papers (Fama and French (1988); Lior Menzly (2004)) found the price-dividend ratio to be an inaccurate proxy for expected dividend growth. Especially, the out-of-sample prediction quality was critically reviewed in these papers.

The literature also deals with the statistical problems that prediction models face. A standard issue is parameter-instability. Changing conditions, no matter of which nature, can have an impact and lead to time-varying coefficients that can represent a significant problem in the specifications of prediction models. Further, the linearity condition of the standard Kalman Filter used in state-space model parameter estimation causes criticism. Also, the log-linear approximation used in the derivation of the present-value model might lead to inaccurate results (Van Binsbergen and Koijen (2011) address these problems and use an unscented Kalman Filter, which can deal with non-linear equations).

To a large extent, prediction literature is based on North American stock data, but there are also multiple publications that examine the existence of predictability in other markets. The studies by Lund and Engsted (1996) and Ang and Bekaert (2006) analyze the interdependence between returns and dividends in the Danish, German, Swedish, UK and multiple other stock markets. When it comes to the present-value model literature, I find most published papers concentrate on US stock market data which is mainly provided by the database of the Center for Research in Security Prices (Campbell and Shiller, 1988a; Cochrane, 2007; Koijen and Van Nieuwerburgh, 2011; Van Binsbergen and Koijen, 2010). There were only a few attempts that apply the present-value state-space representation in other markets, which motivates the application on the German stock exchange.

3 Theory & Research Approach

To detect predictability in the German stock market, I introduce the famous present-value identity by Campbell and Shiller (1988a). According to this identity, variation in the price-dividend ratio implies forecast ability of either returns or dividend growth rates. If the price/dividend ratio is high, either dividends must rise or prices must decline to maintain the identity. Based on this concept, I construct a linear system which imposes model specifications for expected dividend growth rates and expected returns. Then, I present and motivate a state-space representation for the derived system as done in Van Binsbergen and Koijen (2010). Based on this model, I describe the Kalman Filter recursion, for which I provide an extensive derivation in Appendix B. This filter set-up generates estimates for the state variables of the model and further provides us with a likelihood-function for the model parameter estimation. Ultimately, I present Simulated Annealing as an optimization algorithm and apply it on the log-likelihood of the system. In this way, I obtain optimal parameters for the system which again provides me with forecasts for returns and dividend growth rates.

Building upon this approach, I evaluate the parameter estimates and analyze the goodness of fit to the CDAX time series. To examine the results, I calculate R-squared measures and compare them to simple benchmark models. Finally, I conduct hypothesis tests concerning the predictability in the German Stock market and discuss the validity of my results.

All computations are executed via an extensive R-script which is provided in addition to this thesis. To ensure that the model-specifications and the R-code are correct, I applied the algorithm on the same time period of the CRSP-data set as done in the paper of Van Binsbergen and Koijen (2010) (see also Section 4). There are small deviations in my summary statistics of CRSP time series compared to the ones reported in the paper. However, the resulting parameter estimates (with one exception, see Appendix D), plots and R-squared values are basically equal to the

ones of Binsbergen and Kojien, which is why I can assume the correctness of my model and the corresponding computations.

4 Data

In my paper, I want to model and analyze the stock predictability of German stocks on the Frankfurt Stock Exchange using the present-value model. For the approach I need to obtain cum- and ex-dividend price-levels of a representative index. A prominent representative index of the German stock market is given by the DAX. However, the DAX only gives a somewhat limited insight of stock and dividend behavior, since it only includes the 30 largest companies on the exchange. We are looking for a more general result, which should consist of all types of traded stock on the market. A good representative of these stocks is the CDAX. It is a composite stock market index that contains all shares which are traded on the Frankfurt Stock Exchange and fulfill the requirements of General Standard and Prime Standard¹.

Via the application of the present-value identity, I further investigate the relationship between the price-dividend ratio, expected stock returns and expected dividend growth. To obtain information about the dividends, it is therefore essential to collect the cum-dividend and ex-dividend prices of the CDAX. These are given by the performance index (total return index), which incorporates changes in the price-levels as well as dividend-returns, and the stock price index, which excludes dividend returns. I obtain data for the longest available time period from the Bloomberg Terminal. Starting in December 1987 and ending in December 2018, I collect a sample of 372 monthly observations.

The analysis of dividend payouts over time confronts us with a common challenge in financial time series analysis. Dividend payouts are heavily centered in certain months of the year and reveal a strong seasonal component which could distort our

¹These are the two main segments including specific transparency requirements at the Frankfurt Stock Exchange.

Table 1: Average Dividend Payouts of the CDAX during the Year (1989-2018)

Month	Average Dividend Payout
January	0.470
February	0.187
March	0.112
April	1.884
May	4.018
June	0.634
July	0.315
August	0.053
September	0.030
October	0.033
November	0.033
December	0.121

results when analyzing dividend-growth predictability. We can visualize this issue by calculating the monthly average of the dividend payouts, see Table 1. Most of the companies listed on the CDAX pay out their dividends in May.

We can avoid the monthly seasonality in the data by aggregating our data set to annual observations. In terms of the dividends, this could be done by simply summing up the monthly values as in Ang and Bekaert (2006). However, by doing so, we neglect the time value of money. For this reason, we should consider reinvestment strategies for our dividends. One strategy, which is often referred to as Cash-Reinvestment, reinvests the received dividends at the risk-free asset. For this purpose, we need to obtain a low-risk bond comparable to the 30-day treasury bills on the US stock market. Unfortunately, the German Government has not been offering short-term bonds over the whole observed time horizon. Therefore, I consider the 10-year German Government Bond yield (Bund Yield) as an adequate proxy for the risk-free rate. Since the monthly quotes of the yield are commonly noted on an annual base, I need to adjust the rates by multiplying the quotes to the power of $1/12$ to obtain a monthly rate.

Having obtained the data, we can subsequently calculate the basic measurements for our model. We define R_t as the cum-dividend and R_t^{ex} as the ex-dividend returns of the CDAX as following:

$$R_t = \frac{P_t + D_t}{P_{t-1}}, \quad (1)$$

$$R_t^{ex} = \frac{P_t}{P_{t-1}} \quad (2)$$

where P_t denotes the ex-dividend CDAX stock price and D_t denotes the paid out dividends at time t . The annualized version of these returns is given by simply compounding the twelve subsequent monthly returns:

$$R_t^* = R_t \cdot R_{t-1} \cdot \dots \cdot R_{t-11}$$

We can obtain the monthly dividends by deducting the ex-dividend return from the cum-dividend return and multiplying the result by the ex-dividend price:

$$D_t = (R_t - R_t^{ex}) * P_{t-1}$$

For the so-called cash-reinvestment, we denote r_t^f as our risk-free rate in the form of the monthly Bund-yield. To achieve annualized dividend values, I consider the method mentioned Koijen and Van Nieuwerburgh (2011): Each monthly dividend is compounded with every single monthly risk-free rate r_t^f until the end of the year. Consequentially, the annualized compounded dividend in month t is given n by:

$$\begin{aligned} D_t^* &= D_{t-11} \cdot (1 + r_{t-11}^f) \cdot (1 + r_{t-10}^f) \cdot \dots \cdot (1 + r_{t-1}^f) \\ &\quad + D_{t-10} \cdot (1 + r_{t-10}^f) \cdot (1 + r_{t-9}^f) \cdot \dots \cdot (1 + r_{t-1}^f) \\ &\quad \dots \\ &\quad + D_{t-11} \cdot (1 + r_{t-1}^f) \\ &\quad + D_t \end{aligned}$$

Using these return and dividend definitions, the price-dividend ratio can be computed as:

$$PD_t = \frac{P_t}{D_t^*}$$

The yearly dividend growth is simply given by:

$$\Delta D_t^* = \frac{D_t^*}{D_{t-12}^*}.$$

Since we are working in a log-linearized environment (as will be explained in Section 5.1), we compute the logarithms of our measures R_{t+1} , D_{t+1}/D_t and PD_t :

$$r_{t+1} = \ln(R_{t+1}^*) \quad (3)$$

$$\Delta d_{t+1} = \ln(\Delta D_t^*) \quad (4)$$

$$pd_t = \ln(PD_t). \quad (5)$$

Note, that throughout my thesis I use lower case letters to denote the log-representations of the variables. For the computations in Section 5 and 6, I consider the annualized returns and dividend growth rates recorded in December of each year. In this way, we obtain a set of 30 annual observations considering the period from January to December in each year. The resulting summary statistics of the data can be seen in Table 2.

Table 2: Annual Summary Statistics in the case of Cash-reinvested Dividends (CDAX, 1990-2018)

	Δd_t	r_t	pd_t
Mean	0.0622	0.0586	3.7411
Median	0.0529	0.1073	3.7203
Standard Deviation	0.1801	0.2316	0.3213
Maximum	0.5800	0.3685	4.4261
Minimum	-0.3187	-0.5548	2.9935
No. Observations	30	30	30

With only 30 observations, the size of my primary data set is quite limited, which might have a negative impact on the validity of the results. Therefore, I create a second data set in which I record observations semi-annually. The dividends re-

main annualized, while the formulas of the dividend growth and returns are adjusted according to the method explained in Ang and Bekaert (2006). In this case, we compute the returns, dividend growth and price-dividend ratio using the equations:

$$R_t^{semi} = R_t \cdot R_{t-1} \cdot \dots \cdot R_{t-5} \quad (6)$$

$$\Delta D_t^{semi} = \frac{D_t^*}{D_{t-6}^*} \quad (7)$$

$$PD_t^{semi} = \frac{P_t}{D_t^*} \quad (8)$$

The corresponding summary statistics are shown in Table 3. While the mean and median of dividend growth and returns are as expected smaller for the semi-annual series, the statistics of the price-dividend ratio are very similar.

Table 3: Semi-Annual Summary Statistics in the case of Cash-reinvested Dividends (CDAX, 1989-2018)

	Δd_t	r_t	pd_t
Mean	0.0311	0.0293	3.7450
Median	0.0169	0.0582	3.7220
Standard Deviation	0.1134	0.1463	0.3075
Maximum	0.3856	0.2933	4.4261
Minimum	-0.2000	-0.3925	2.9935
No. Observations	59	59	59

Further, I obtain cum- and ex-dividend returns and price levels for the aggregated American stock market provided by the CRSP database from 1946–2018. This data set is used in multiple publications that apply the Present-Value model in State-Space form, including the papers of Van Binsbergen and Koijen (2010) and Piatti and Trojani (2017). To validate the functionality of my model I have compared my computations to the ones of Van Binsbergen and Koijen. For each of the following computations, I have attached the results based on the CRSP data set in Appendix D.

5 Methodology

To construct the model as presented by Van Binsbergen and Koijen (2010), I start by deriving the present-value identity. Then, I set up the state-space representation and derive the Kalman Filter recursion. Based on the filtering process, I obtain a likelihood-function which is subsequently maximized by the introduced Simulated Annealing Algorithm. Ultimately, I obtain estimates for the parameters of our model, which are going to be analyzed in Section 6. For simplicity, I denote annualized dividends D_t^* as D_t and the compounded annualized returns R_t^* as R_t in the following derivation, where t refers to the year of the observation (or a six-month period in terms of the semi-annual data set).

5.1 The Present-Value Identity

The model used in my thesis is based on the present-value identity by Campbell and Shiller (1988b). In this section, I derive the log-linearized return relation of the price-dividend ratio and subsequently obtain the present-value equation, which is essential for my approach. For transparency, I only describe the main steps of the derivation while a more detailed derivation can be found in Appendix A.

I start by defining a simple return-identity using the cum-dividend return definition 1:

$$1 = R_{t+1}^{-1} \cdot R_{t+1} = R_{t+1}^{-1} \cdot \frac{P_{t+1} + D_{t+1}}{P_t}$$

Multiplying by P_t/D_t results in

$$\begin{aligned} \frac{P_t}{D_t} &= R_{t+1}^{-1} \cdot \frac{P_{t+1} + D_{t+1}}{P_t} \cdot \frac{P_t}{D_t} \\ &= R_{t+1}^{-1} \cdot \left(1 + \frac{P_{t+1}}{D_t}\right) \cdot \frac{D_{t+1}}{D_t} \end{aligned}$$

We can now take logs on both sides to obtain a log-linearized expression of the price-dividend ratio pd_t . By using the property $P_t/D_t = \exp[\ln(P_t/D_t)] = \exp(pd_t)$ and inserting the notations 3–5 we obtain:

$$pd_t = -r_{t+1} + \Delta d_{t+1} + \ln[1 + \exp(pd_{t+1})]$$

The last term, $\ln[1 + \exp(pd_{t+1})]$, can be treated with a first-order Taylor Expansion (see equation A2) around a point $\overline{pd} = E[pd_t]$ (typically the historical mean) to get the following log-linearized approximation of the price-dividend ratio:

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

where κ and ρ are defined by

$$\kappa = \ln[1 + \exp(\overline{pd})] - \rho \overline{pd} \quad \text{and} \quad \rho = \frac{\exp(\overline{pd})}{\exp(1 + \overline{pd})} \quad (9)$$

The use of this linearization might contain an approximation error which could lead to biased results in the ultimate present-value model (further discussed in the validity check in Section 6.4). However, the interpretation of this equation corresponds to common economic intuition: Given a fixed price-dividend ratio at time t , a higher dividend growth rate at $t + 1$ implies higher future dividend payments, which again have a positive effect on future returns. If the price at time t is high and correspondingly the price-dividend ratio is high, we expect the future returns to be lower. On the other hand, a higher price-dividend ratio at $t + 1$ should come along with higher returns in this period.

We can iterate this equation forward in time by gradually substituting for pd_{t+i} :

$$\begin{aligned} pd_t &= \kappa + \rho pd_{t+1} + \Delta d_{t+1} - r_{t+1} \\ &= \kappa + \rho(\kappa + \rho pd_{t+2} + \Delta d_{t+2} - r_{t+2}) + \Delta d_{t+1} - r_{t+1} \\ &= \kappa + \rho\kappa + \rho^2 pd_{t+2} + (\Delta d_{t+1} - r_{t+1}) + \rho(\Delta d_{t+2} - r_{t+2}) \\ &= \dots \end{aligned}$$

$$= \sum_{j=0}^{\infty} \rho^j \kappa + \rho^{\infty} p d_{\infty} + \sum_{j=1}^{\infty} \rho^{j-1} (\Delta d_{t+j} - r_{t+j})$$

Since $\rho < 1$, by definition 9, we can assume that:

$$\rho^{\infty} p d_{\infty} = \lim_{j \rightarrow \infty} \rho^j p d_j = 0$$

If we additionally consider the properties of an infinite geometric series (see equation A3), we can rewrite this equation:

$$p d_t = \frac{\kappa}{1 - \rho} + \sum_{j=1}^{\infty} \rho^{j-1} (\Delta d_{t+j} - r_{t+j})$$

Now, we can take expectations conditional upon time to define a relation between price-dividend ratio, expected returns and expected dividend growth. Because this equation holds ex-ante and ex-post, the expectation operator can be added on the right-hand side:

$$\begin{aligned} p d_t &= E_t \left[\frac{\kappa}{1 - \rho} + \sum_{j=1}^{\infty} \rho^{j-1} (\Delta d_{t+j} - r_{t+j}) \right] \\ &= \frac{\kappa}{1 - \rho} + \sum_{j=1}^{\infty} \rho^{j-1} E_t [\Delta d_{t+j} - r_{t+j}] \end{aligned}$$

Before we continue, we have to define the time series properties of expected returns and expected dividend growth rates. For the sake of the derivation, we follow the common assumption, that both variables follow AR(1)-processes (see also Pástor and Stambaugh (2009) or Van Binsbergen and Koijen (2010)). Consequentially, we define:

$$E_t[r_{t+1}] = \mu_{t+1} = \delta_0 + \delta_1(\mu_t - \delta_0) + \epsilon_{t+1}^{\mu} \quad (10)$$

$$E_t[\Delta d_{t+1}] = g_{t+1} = \gamma_0 + \gamma_1(g_t - \gamma_0) + \epsilon_{t+1}^g \quad (11)$$

for expected returns and expected dividend-growth, where the error terms ϵ_{t+1}^μ and ϵ_{t+1}^g are assumed to have zero mean. We can now substitute μ_t and g_t into the equation:

$$\begin{aligned} pd_t &= \frac{\kappa}{1-\rho} + \sum_{j=1}^{\infty} \rho^{j-1} E_t [g_{t+j-1} - \mu_{t+j-1}] \\ &= \frac{\kappa}{1-\rho} + \sum_{j=0}^{\infty} \rho^j E_t [g_{t+j} - \mu_{t+j}] \end{aligned}$$

and make use of the AR(1)-properties of the expected returns and expected dividend growth (see equation A4):

$$\begin{aligned} pd_t &= \frac{\kappa}{1-\rho} + \sum_{j=0}^{\infty} \rho^j E_t [g_{t+j} - \mu_{t+j}] \\ &= \frac{\kappa}{1-\rho} + \sum_{j=0}^{\infty} \rho^j [\gamma_0 + \gamma_1(g_{t+j-1} - \gamma_0) - \delta_0 - \delta_1(\mu_{t+j-1} - \delta_0)] \end{aligned}$$

Note that the error-terms of the AR-processes have zero mean and can be omitted after taking expectations. We can now substitute for g_t and μ_t and iterate the two terms forward considering the following property of the AR-process (see also equation A4):

$$\Leftrightarrow E[\mu_{t+j}] = \delta_0 + \delta_1^j(\mu_t - \delta_0)$$

The same can be applied to the AR-process of the expected dividend growth.

Thereby we reach:

$$\begin{aligned} pd_t &= \frac{\kappa}{1-\rho} + \sum_{j=0}^{\infty} \rho^j [\gamma_0 + \gamma_1(g_{t+j-1} - \gamma_0) - \delta_0 - \delta_1(\mu_{t+j-1} - \delta_0)] \\ &= \frac{\kappa}{1-\rho} + \sum_{j=0}^{\infty} \rho^j [\gamma_0 + \gamma_1^j(g_t - \gamma_0) - \delta_0 - \delta_1^j(\mu_t - \delta_0)] \\ &= \frac{\kappa}{1-\rho} + \sum_{j=0}^{\infty} \rho^j(\gamma_0 - \delta_0) + \sum_{j=0}^{\infty} \rho^j [\gamma_1^j(g_t - \gamma_0) - \delta_1^j(\mu_t - \delta_0)] \end{aligned}$$

We can now make use of the properties of infinite geometric series again and reach:

$$pd_t = \frac{\kappa}{1-\rho} + \frac{\gamma_0 - \delta_0}{1-\rho} + \frac{g_t - \gamma_0}{1-\rho\gamma_1} - \frac{\mu_t - \delta_0}{1-\rho\delta_1}.$$

From here we can form the final present-value equation which links the price-dividend ratio, the expected returns and the expected dividend growth:

$$pd_t = A - B_1(\mu_t - \delta_0) + B_2(g_t - \gamma_0) \quad (12)$$

with

$$A = \frac{\kappa}{1 - \rho} + \frac{\gamma_0 - \delta_0}{1 - \rho} \quad (13)$$

$$B_1 = \frac{1}{1 - \rho\delta_1} \quad (14)$$

$$B_2 = \frac{1}{1 - \rho\gamma_1} \quad (15)$$

We can see that the log price-dividend ration can be expressed as a linear combination of the expected returns (μ_t) and the expected dividend growth rates (g_t). The impact of these two latent variables on the price-dividend ratio depends on their relative persistence represented by the two constants δ_1 and γ_1 .

5.2 The Present-Value Model

Having derived the identity, we can sum up the base for the present-value model. First, let's recall that expected dividend growth rates (g_t) and expected stock returns (μ_t) are following AR(1) processes:

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

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

where

$$\mu_t = E_t[r_{t+1}]$$

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

The dividend growth at $t + 1$ can be modeled as its expected value plus an error term. Thus, using the latter expression, we obtain:

$$\Delta d_{t+1} = g_t + \epsilon_{t+1}^d$$

Lastly, the present-value identity is given by

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

with

$$A = \frac{\kappa}{1 - \rho} + \frac{\gamma_0 - \delta_0}{1 - \rho} \quad (16)$$

$$B_1 = \frac{1}{1 - \rho\delta_1} \quad (17)$$

$$B_2 = \frac{1}{1 - \rho\gamma_1} \quad (18)$$

Through these equations, we obtain a dynamic linear system. The three defined processes for expected returns, expected dividend growth rates and realized growth rates contain three error-terms which can be inter-correlated: ϵ_{t+1}^μ , ϵ_{t+1}^g and ϵ_{t+1}^d . We assume that each of these is independent and identically distributed (i.i.d.) over time. They have zero mean and the following covariance matrix:

$$\Sigma = \text{var} \left(\begin{bmatrix} \epsilon_{t+1}^g \\ \epsilon_{t+1}^\mu \\ \epsilon_{t+1}^d \end{bmatrix} \right) = \begin{pmatrix} \sigma_g^2 & \sigma_{g\mu} & \sigma_{gd} \\ \sigma_{g\mu} & \sigma_\mu^2 & \sigma_{\mu d} \\ \sigma_{gd} & \sigma_{\mu d} & \sigma_d^2 \end{pmatrix}.$$

This set of equations provides us with the base for a state-space representation of the present-value model. However, we can further transform the system to obtain a more convenient form of the desired model. Therefore, I define:

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

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

as the de-measured state variables following the corresponding AR(1)-process:

$$\hat{\mu}_{t+1} = \delta_1 \hat{\mu}_t + \epsilon_{t+1}^\mu \quad (19)$$

$$\hat{g}_{t+1} = \gamma_1 \hat{g}_t + \epsilon_{t+1}^g \quad (20)$$

These two expressions represent the so-called transition equations of the latent variables (also called state-equations, as described in Section 5.3). Consequentially, dividend growth and price-dividend ratio are given by:

$$\Delta d_{t+1} = \gamma_0 + \hat{g}_t + \epsilon_{t+1}^d$$

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

with A , B_1 , and B_2 as defined in the equations 13–15. Since there is no error term in our price-dividend ratio equation, we can substitute it in one of the transition equations to simplify the model. We start by rearranging:

$$\begin{aligned} pd_t &= A - B_1 \hat{\mu}_t + B_2 \hat{g}_t \\ \Leftrightarrow \hat{\mu}_t &= \frac{1}{B_1} (A + B_2 \hat{g}_t - pd_t) \\ \Leftrightarrow \hat{\mu}_{t+1} &= \frac{1}{B_1} (A + B_2 \hat{g}_{t+1} - pd_{t+1}) \end{aligned}$$

Substituting these terms into the AR(1)-equation of the expected returns (equation 19) results in:

$$\begin{aligned} \hat{\mu}_{t+1} &= \delta_1 \hat{\mu}_t + \epsilon_{t+1}^\mu \\ \Leftrightarrow \frac{1}{B_1} (A + B_2 \hat{g}_{t+1} - pd_{t+1}) &= \delta_1 \frac{1}{B_1} (A + B_2 \hat{g}_t - pd_t) + \epsilon_{t+1}^\mu \\ \Leftrightarrow A + B_2 \hat{g}_{t+1} - pd_{t+1} &= \delta_1 (A + B_2 \hat{g}_t - pd_t) + B_1 \epsilon_{t+1}^\mu \\ \Leftrightarrow pd_{t+1} &= -\delta_1 (A + B_2 \hat{g}_t - pd_t) - B_1 \epsilon_{t+1}^\mu + A + B_2 \hat{g}_{t+1} \\ \Leftrightarrow pd_{t+1} &= -\delta_1 (A + B_2 \hat{g}_t - pd_t) - B_1 \epsilon_{t+1}^\mu + A + B_2 (\gamma_1 \hat{g}_t + \epsilon_{t+1}^g) \\ \Leftrightarrow pd_{t+1} &= (1 - \delta_1) A + B_2 (\gamma_1 - \delta_1) \hat{g}_t + \delta_1 pd_t - B_1 \epsilon_{t+1}^\mu + B_2 \epsilon_{t+1}^g \end{aligned}$$

Ultimately, this leads to the final system for the present-value model under cash-reinvested dividends:

$$\hat{g}_{t+1} = \gamma_1 \hat{g}_t + \epsilon_{t+1}^g, \quad (21)$$

$$\Delta d_{t+1} = \gamma_0 + \hat{g}_t + \epsilon_{t+1}^d, \quad (22)$$

$$pd_{t+1} = (1 - \delta_1)A + B_2(\gamma_1 - \delta_1)\hat{g}_t + \delta_1 pd_t - B_1 \epsilon_{t+1}^\mu + B_2 \epsilon_{t+1}^g. \quad (23)$$

This leaves us with a vector of parameters that needs estimating:

$$\Theta = (\gamma_0, \delta_0, \gamma_1, \delta_1, \sigma_g, \sigma_\mu, \sigma_d, \rho_{g\mu}, \rho_{gd}, \rho_{\mu d}). \quad (24)$$

The process of estimation within a state-space representation is the core of the following sections.

5.3 State Space Models for Time Series Analysis

Introduced in a pioneering paper by Kalman (1960), State Space models (sometimes also called dynamic linear models) can model dynamic systems in which unobserved and observed variables evolve over time and are causally connected with each other. Even though this approach was initially invented for the field of control-engineering, it turned out to be very useful for time series analysis in economics. In this section, I will present the basic form of a linear Gaussian State Space Model as described in Hamilton (1994) or Durbin and Koopman (2012) and relate it to the present-value model.

The state space is a Euclidean Space in which the unobserved variables or states can be described via a vector within this space. In its basic form, we can form the so-called state equation as a vector AR-process:

$$X_{t+1} = FX_t + \Gamma \epsilon_{t+1}^X, \text{ with } \epsilon_{t+1}^X \sim \mathcal{N}(0, \Sigma) \quad (25)$$

where X_t represents the $r \times 1$ state-vector, F is a $r \times r$ matrix, and Γ is in our case a subset of the Identity matrix and therefore also called selection matrix. The error

terms contained in ϵ_{t+1}^X are assumed to be independent and identically distributed (i.i.d.) over time (serial independence), have zero mean and a constant covariance matrix Σ .

We cannot model the state in a classic manner like in a simple least squares regression since the states are assumed to be unobservable. However, in the State Space approach, we introduce a vector of the observable variables Y_t which is in itself a linear transformed version of the state. Based on these measurements, we can infer the values of the state variables, which are often called latent variables because of this attribute. It follows the basic form of an observation (or measurement) equation:

$$Y_t = C' z_t + H' X_t \quad (26)$$

where Y_t and z_t are vectors of the dimension $(n \times 1)$ and $(k \times 1)$. C' and H represent predetermined matrices of the dimensions $(n \times k)$ and $(n \times r)$. $C' z_t$ is the $p \times m$ observation matrix. It can include any external variables or previous values of Y_t that have an impact on the measurement.

5.3.1 The Present-Value State-Space Model

State space models can generate estimates for unobservable variables based on their relation to observable variables. I am applying this approach on the dynamics between the unobserved state variables, in form of expected dividend growth-rates g_t and expected returns μ_t , and the causally related measurements of realized dividend growth rates Δd_t and log-price-dividend ratios pd_t as defined in the equations 21–23. The corresponding state-space model is given by:

$$\begin{aligned} \hat{g}_{t+1} &= \gamma_1 \hat{g}_t + \epsilon_{t+1}^g \\ \Delta d_{t+1} &= \gamma_0 + \hat{g}_t + \epsilon_{t+1}^d \\ pd_{t+1} &= (1 - \delta_1)A + B_2(\gamma_1 - \delta_1)\hat{g}_t + \delta_1 pd_t - B_1 \epsilon_{t+1}^\mu + B_2 \epsilon_{t+1}^g \end{aligned}$$

We can reformulate the model in a standardized State-Space form by defining:

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

as our state equation with the state vector:

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

and

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

The vector of the error terms is given by:

$$\epsilon_{t+1}^X = \begin{bmatrix} \epsilon_{t+1}^d \\ \epsilon_{t+1}^g \\ \epsilon_{t+1}^\mu \end{bmatrix}$$

which is assumed to be serial independent over time with zero-means and a covariance matrix:

$$\Sigma = var \left(\begin{bmatrix} \epsilon_{t+1}^g \\ \epsilon_{t+1}^\mu \\ \epsilon_{t+1}^d \end{bmatrix} \right) = \begin{pmatrix} \sigma_d^2 & \sigma_{dg} & \sigma_{d\mu} \\ \sigma_{dg} & \sigma_g^2 & \sigma_{g\mu} \\ \sigma_{d\mu} & \sigma_{g\mu} & \sigma_\mu^2 \end{pmatrix}.$$

The vector for the observed measurements is given by $Y_t = (\Delta d_t, pd_t)$, and we can define the observation equation based on the present value model under cash-reinvested dividends as following:

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

where

$$M_0 = \begin{bmatrix} \gamma_0 \\ (1 - \delta_1)A \end{bmatrix}, M_1 = \begin{bmatrix} 0 & 0 \\ 0 & \delta_1 \end{bmatrix}$$

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

Note that the product $M_0 + M_1 Y_{t-1}$ is the representative term for the observation matrix $C' z_t$ in equation 26.

5.4 The Kalman Filter

In time series analysis filtering describes the process of treating data by removing unwanted components such as noise. In finance, one of the most applied filtering algorithms is the Kalman Filter, which is set up on the base of a state-space model. It proved to be very useful for noisy observations as we often find them in many economic time series systems. At each step in time, the Kalman filter is able to generate optimal estimates for the unobservable state parameters of a system (in our case the expected dividend growth). Further, it can compute predictions of the state variables. The generated estimates are recursively adjusted at each step in time based on the incoming observable measurements. In this section, I sum up the Kalman-filtering process for the derived state-space model under cash-reinvested dividends. A detailed derivation of each step in the process including the made assumptions can be found in Appendix B.

The Kalman filter loops through the observation set and generate estimates at each time step according to the relations defined in the State-Space model. Before the start of the recursion, we need to define an initial estimate of the state vector X_t and its mean squared error (MSE) P_t . There are several ways to determine these values, which are not based on any observations. If the applicant of the model has a clue what the actual value of the state is, he or she could make an educated guess as a starting value. However, the more general approach, which we will apply,

is to assume that the initial values are given by the unconditional mean and the unconditional covariance matrix of the state. For our model, the initial state is given by:

$$X_{0|0} = E[X_1] = E[X_t] = 0_{rx1} \quad (27)$$

and the corresponding covariance matrix is:

$$P_{0|0} = E[(X_t - E[X_t])(X_t - E[X_t])'] = E[X_t X_t'] \quad (28)$$

which can be solved via the formula:

$$vec(P_{0|0}) = [I_{r^2} - (F \otimes F)]^{-1} vec(\Gamma \Sigma \Gamma') \quad (29)$$

where $(F \otimes F)$ denotes the Kronecker product (see Appendix C). Having obtained the initial values, we can start the Kalman Filter recursion. The first step is to generate a forecast of the state at time t based on its previous updated version at time $t - 1$ (on the first recursion, this version is given by $X_{0|0}$). Based on the definition of the state-equation, the forecast and the corresponding MSE are given by:

$$\begin{aligned} \hat{X}_{t|t-1} &= F X_{t-1|t-1} \\ P_{t|t-1} &= F P_{t-1|t-1} F' + \Gamma \Sigma \Gamma' \end{aligned}$$

In the next step, we compute a forecast of the measurement Y_t based on $\hat{X}_{t+1|t}$ and then compare it to the actual value of Y_t by calculating the corresponding forecasting error η_t . The forecasts for the observations are then given by:

$$\hat{Y}_{t|t-1} = M_0 + M_1 Y_{t-1} + M_2 \hat{X}_{t|t-1}$$

The error is computed via:

$$\eta_t = Y_t - M_0 - M_1 Y_{t-1} - M_2 \hat{X}_{t|t-1}$$

We can further define the MSE of η_t , which will be essential for the log-likelihood function presented in section 5.4.1. It is denoted as S_t and is obtained by calculating:

$$S_t = E[(Y_t - \hat{Y}_{t|t-1})(Y_t - \hat{Y}_{t|t-1})'] = M_2 P_{t|t-1} M_2'$$

Lastly, the last step of the recursion updates our estimates of the state and the corresponding covariance matrix according to the obtained forecasting error and the so-called Kalman-Gain Matrix, which is defined as:

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

Lastly, we compute the updated state-vector and its covariance matrix by:

$$\begin{aligned}\hat{X}_{t|t} &= \hat{X}_{t|t-1} + K_t \eta_t \\ P_{t|t} &= (I_4 - K_t M_2) P_{t|t-1}.\end{aligned}$$

If we re-substitute the terms

$$\Gamma \epsilon_{t+1}^X = v_{t+1}, \text{ with } E[v_t v_\tau'] = \begin{cases} Q = \Gamma \Sigma \Gamma' \text{ for } t = \tau \\ 0 \text{ otherwise} \end{cases},$$

$$C' z_t = M_0 + M_1 Y_{t-1},$$

we end up with the same process as described in Van Binsbergen and Koijen (2010):

$$X_{0|0} = E[X_0] = 0_r$$

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

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

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

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

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

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

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

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

where r describes the size of the state vector.

5.4.1 Maximum Likelihood Estimation of the Model Parameters

In Appendix B, we derive the forecasts $\hat{X}_{t|t-1}$ and $\hat{Y}_{t|t-1}$ in the sense of linear projections. They therefore represent optimal linear forecasts conditional on the information contained in the previous observations $\Upsilon_{t-1} = Y_{t-1}, \dots, Y_1$ in any case. Furthermore, the errors in our state-space model are assumed to be normally distributed, which makes it possible to make an even stronger statement. Under these circumstance the forecasts of our Kalman Filter are optimal in the light of any function of (Υ_{t-1}) . It also implies that Y_t , conditional on Υ_{t-1} , is normally distributed with the mean $\hat{Y}_{t+1|t}$ and variance S_t :

$$Y_t | z_t, \Upsilon_{t-1} \sim N(\hat{Y}_{t+1|t}, S_t)$$

Consequently, the distribution can be described by the Gaussian density function:

$$f_{Y_t | \Upsilon_{t-1}}(\Delta d_t, p d_t) = \frac{\exp \left[-\frac{1}{2} (Y_t - \hat{Y}_{t+1|t})' S_t^{-1} (Y_t - \hat{Y}_{t+1|t}) \right]}{\sqrt{(2\pi)^2 |S_t|}}$$

$$= \frac{\exp \left(-\frac{1}{2} \eta_t' S_t^{-1} \eta_t \right)}{\sqrt{(2\pi)^2 |S_t|}}$$

We can now easily derive the log-likelihood function:

$$l_t = \ln(f_{Y_t | z_t, \Upsilon_{t-1}}(\Delta d_t, p d_t))$$

$$= \ln \left[\exp \left(-\frac{1}{2} \eta_t' S_t^{-1} \eta_t \right) \right] - \ln \left(\sqrt{(2\pi)^2 |S_t|} \right)$$

$$= -\frac{1}{2} \eta_t' S_t^{-1} \eta_t - \frac{1}{2} \ln \left((2\pi)^2 |S_t| \right)$$

$$= -\frac{1}{2} \left[\eta_t' S_t^{-1} \eta_t + \ln(|S_t|) \right] - \ln(2\pi)$$

We can omit the last term and the constant factor since they will not affect the optimal solution which maximizes the likelihood. Ultimately, we obtain the log-likelihood function of the Kalman Filter:

$$l_t = -\eta_t' S_t^{-1} \eta_t - \ln(|S_t|)$$

We want to choose the set of parameters Θ (see definition 24) which maximizes the likelihood over the whole series of T observations of Y_t . Therefore, we aim for maximization of the aggregated likelihood function, which is given by:

$$L = - \sum_{t=1}^T \ln(|S_t|) - \sum_{t=1}^T \eta_t' S_t^{-1} \eta_t$$

We subsequently try to maximize this equation via the Simulated Annealing algorithm described in Section 5.6.

5.5 Estimation Restrictions

Before getting started with the optimization of the likelihood function, a few constraints regarding the values of the estimation parameters need to be imposed. First, the Kalman filter comes along with an identity issue, if we do not place any restrictions on F , Q , C and M_2 . If the parameters of our state-space model are unidentified, there is more than one set of parameter values that could result in the same likelihood-values. Consequentially, we would not be able to find the optimal parameter set for our present-value model. This is why we predetermine the correlation between the error terms for realized dividend growth and for the expected dividend growth to be zero:

$$\rho_{dg} = 0$$

Thereby, we make sure that all the parameters in the covariance matrix of the error-terms are identified.

Second, we need to set upper and lower boundaries for the rest of the parameters to be estimated. These will make sure that the covariance matrices in the model

stay positive definite and that the AR-process of the state variable is covariance stationary (which is also essential for the derivation of the Kalman Filter). We ensure stationarity by defining:

$$|\gamma_1| < 1 \text{ and } |\delta_1| < 1$$

In the case of the cash-reinvestment model the covariance matrix of the shocks stays positive definite by constraining the standard deviations of the shocks:

$$\sigma_g, \sigma_\mu, \sigma_d > 0$$

The correlation parameters are bound between -1 and 1:

$$-1 < \rho_{g\mu}, \rho_{\mu d} < 1$$

5.6 Simulated Annealing

Having its origin in the field of thermodynamics (which will be visible in the naming of the parameters), Simulated Annealing (SA) represents a numerical optimization algorithm that proved to be particularly useful when searching for the optimal parameters of complex models in economics. Depending on the properties and the complexity of an optimization problem, other conventional algorithms sometimes struggle to find a global maximum. Often, these implementations cannot distinguish between local and global maxima, use too many calculation steps (and thus computation time), converge to infinite parameter values or get stuck and do not find any solution at all. Furthermore, the choice of the right starting values often plays a role in the quality of the results, which can, for itself, create another problem, which has to be solved first.

SA can be superior to many other algorithms considering these aspects. SA searches for the global maximum across the whole surface of the treated function. In the process, it moves both uphill and downhill, not getting caught at a single local

maximum. Because of this attribute, the algorithm is, to a large extent, independent of the choice of starting values for the parameters. Further, SA does not have strict requirements for the function. It can handle functions that have ridges or plateaus, are not continuous or are not defined for specific parameter values.

We find some of these aspects in our optimization problem in the form of the maximization of the likelihood implied by the Kalman filter. The function shows holes in the surface, it has restricted parameters bound by certain intervals (e.g. the correlation parameters), and it is, in general, a high-dimensional rather complex problem to solve. Therefore, SA represents a good choice to find optimal parameters for our present value model. In this paper, I use a slightly altered version of the simulated annealing algorithm in the spirit of the approach of William L. Goffee and Rogers (1994) that builds upon the SA implementation by Corana et al. (1987). I describe the algorithm concerning our likelihood maximization in more detail in section 5.6.

We start by defining $L(\Theta)$ as our likelihood function with the parameter vector Θ . We predefine the initial objects of the Kalman Filter, X_0 and P_0 , and collect the time series of observations Y_t . We further impose boundaries for the estimation parameters to limit the surface the algorithm has to search through according to the definitions in Section 5.5. Furthermore, SA requires the presetting of the annealing parameters for the initialization. These parameters decide on the accuracy and the speed of the computation. So, before starting the algorithm, we need to choose the initial values of the parameter vector Θ_0 and store the corresponding function value for this set. We define an initial step vector V_0 , its adjustment vector C and an initial temperature value T_0 . For the termination criterion, I set a threshold value ϵ and a number N_ϵ . Further, we need to set the numbers N_S and N_T , which limit the number of function evaluations before the step vector is adjusted, the termination criteria is controlled or the temperature is reduced. Ultimately, we need to set a factor r_T for the temperature reduction after every N_T loops and define a lower bound for the temperature as additional termination criteria.

Having defined these objects, we can start the main loop of the algorithm. In the first step, we set a new parameter vector Θ' by randomly altering each element of the initial vector via the equation:

$$\theta'_i = \theta_i + r * v_i$$

where θ'_i , θ_i , and v_i are elements of the respective vectors Θ' , Θ , and V , and r is a uniformly distributed random on the interval $[-1, 1]$. If the generated parameter value lies outside the predefined boundaries, we repeat the previous step as long as the new parameter is valid. In the next step, the corresponding new function value $L'(\Theta')$ is computed. If $L'(\Theta')$ is larger than the previous value $L(\Theta)$, we accept the new parameters and set Θ to Θ' (the algorithm is moving “uphill” in the surface). Further, if $L'(\Theta')$ is larger than any other previously found function value, we record the value and the parameter vector as our new optimum in $L_{opt}(\Theta_{opt})$ and Θ_{opt} . On the other hand, if $L'(\Theta')$ is lower than $L(\Theta)$, we apply the so-called Metropolis criteria to decide whether the new point is accepted or rejected. The criterion uses the probability measure:

$$p = \exp\left(\frac{L' - L}{T}\right)$$

and compares it to an uniformly distributed random number p' on the interval $[0, 1]$. The parameter vector is accepted if p is larger than p' . If that is the case, the algorithm moves “downhill” to the smaller function value and Θ becomes Θ' . Otherwise, the new parameters are rejected. Looking at the equation for p , we can see that two terms affect the probability of rejection for lower function values. First, the larger the difference between the new and old function value, the more likely is the rejection. Second, the lower the temperature value, the higher the probability of rejection. This implies, that with every reduction of the temperature, the probability of a rejection rises.

After N_S loops through this procedure, the step-vector V is adjusted so that in the next N_S loops fewer parameter moves will be accepted. The step-length elements v_i can be quite large in the beginning so that the SA algorithm can build up a rough overview of the function surface. They should then gradually decrease to restrict the search for a maximum on the most promising areas. Corana et al. (1987) introduce a criterion for this purpose that reduces the number of accepted moves by about 50%. It alters every element of V , and thereby the step length in the parameter alteration, in the following way:

$$\begin{aligned}
 v'_i &= v_i \left(1 + c_i \frac{n_i/N_S - 0.6}{0.4} \right) & \text{if} & \quad \frac{n_i}{N_S} > 0.6 \\
 v'_i &= \frac{v_i}{1 + c_i \frac{0.4 - n_i/N_S}{0.4}} & \text{if} & \quad \frac{n_i}{N_S} < 0.4 \\
 v'_i &= v_i & \text{else} &
 \end{aligned}$$

where n_i is the number of alterations of the i 'th parameter that led to acceptance during the last N_S evaluations. The use of the ratio n_i/N_S implies that if more than 60% of the points are accepted for θ_i , then the corresponding v_i increases leading to more rejections in the ongoing process. On the contrary, if the percentage is less than 40% the element of V decreases leading to fewer rejections (for a given temperature level). It is very useful to analyze the values inside the step-vector at the end of an optimization run. The values are supposed to be very small at the end of a cycle since this implies that the algorithm focused on a minor area and is not jumping over considerable distances in the function's surface.

After N_T of these adjustments, the temperature T is reduced by the factor r_T :

$$T' = r_T T$$

As previously described, a lower temperature leads to more parameter rejections, which makes downhill moves in the surface less likely. This leads to lower n_i values, which then leads to smaller step lengths represented by V . As a result, the

algorithm is continuously focusing on a smaller and more promising area on the surface.

The algorithm stops as soon as the predefined lower temperature boundary is breached or the termination criterion is met. The criterion consists of two parts. First, the difference between the current loop's largest function value $L(\Theta)$ and the optimal value $L_{opt}(\Theta)$ has to be smaller than ϵ . Is this the case, the second criterion is checked: Before each temperature reduction, the highest function value out of the current cycle is recorded and stored in a vector. Then, each of the last N_ϵ of these recorded values is compared to the largest function value of the current cycle. If the differences between each of these values are smaller than ϵ , the criterion is met and the SA algorithm terminates. Otherwise, the temperature is reduced again and the cycle starts all over again (as long as the temperature boundary is not breached). In this way, SA tries to ensure that a global and not a local maximum was reached.

Note that R provides several filtering and optimization function (e.g. "Kalman-Forecast()," "nlm()," or "optim()"), which I tried to apply on the Kalman filter estimation. However, none of the used functions could reach the optimal parameters reported in Van Binsbergen and Koijen (2010), which is why I decided to program the algorithm myself. Ultimately, I managed to obtain the optimal parameter values of van Binsbergen based on the CRSP dataset which validates the applied R-script attached to this thesis.

6 Results

In this section, I am going to analyze the obtained estimates and the performance of the final present-value model. I evaluate the goodness of fit via R^2 -measures and compare it to the ones of simple predictive regressions. Then, I test for the presence of predictability with Likelihood-Ratio tests.

6.1 Estimation Results

After maximizing the likelihood via Simulated Annealing, I obtain the optimal estimates of the model parameters, which are shown in Table 4.

The values of the model parameters δ_0 and γ_0 express the unconditional means of the latent variables. For the annual data, we estimate an unconditional expected return of 6.1%, while for the semi-annual data, we obtain a mean of 5.2%. This is quite surprising considering the considerable difference between the mean of the log-returns in the summary statistics. The unconditional expected dividend growth rate is given 3.7% and 2.7% respectively. Furthermore, we find returns to be highly persistent with a δ_1 value of 0.82 in the annual data and even 0.94 in the semi-annual data set. This is consistent with the results of Van Binsbergen and Kojien (2010) for the CRSP data set. Comparing the estimates of the annual series of the CDAX with the estimates based on the CRSP data set (Appendix D2), we can see that there is a higher persistence in the returns on the American stock market, while the persistence of the expected dividend growth is higher in the German stock market. However, in both series, expected dividend-growth rates are far less persistent as returns. In Section 6.3 we will further test for persistence via a likelihood-ratio test. Looking at the parameters of the covariance matrix of the innovations in the model, we detect highly negative values for the correlations between expected dividend growth and expected returns. This is very surprising since the intuition of the present-value equation would suggest a positive correlation, which is also given in

the CRSP data. In general, the estimates of correlation parameters in the present-value model seem to be quite volatile. If the other parameters are fixed, changes in the correlation parameter do not result in large jumps of the likelihood value. Van Binsbergen and Kojen (2010) come to a similar conclusion as they calculate bootstrapped standard errors and analyze the finite-sample properties of the parameters for the CRSP series (which in addition has more observations than our CDAX data set).

Table 4: Maximum-Likelihood Parameter - Estimates (CDAX, 1990–2018)

Maximum Likelihood Estimates		
Parameter	Estimate (Annual Data)	Estimate (Semi-Annual Data)
γ_0	0.037	0.027
δ_0	0.061	0.052
γ_1	0.359	0.543
δ_1	0.818	0.942
σ_g	0.081	0.041
σ_μ	0.024	0.007
σ_d	0.125	0.098
$\rho_{g\mu}$	-0.939	-0.969
$\rho_{\mu d}$	0.796	0.959
Log-Likelihood-Value	146.106	372.246
Implied Present-Value Model Parameters		
A	3.731	3.693
B_1	4.971	12.614
B_2	1.540	2.130
ρ	0.977	0.977

6.2 In-Sample Performance

Having obtained the estimates of the present-value state-space model, we can now analyze its performance. If the model works well with the analyzed time series, it should be able to generate accurate estimates of dividend-growth and returns. In this section, I examine the goodness of fit of the model via the calculation of in-

sample R^2 -measures for returns and dividend growth rates. I further compare the quality of the estimates compared to simple predictive regression models. For the evaluation of the fit, I calculate the R^2 -measure as described in Harvey (1990). For dividend-growth rates and returns, this measure is defined as:

$$R_{Div}^2 = 1 - \frac{v\hat{a}r(\Delta d_{t+1} - g_t^F)}{v\hat{a}r(\Delta d_{t+1})} \quad (30)$$

$$R_{Ret}^2 = 1 - \frac{v\hat{a}r(r_{t+1} - \mu_t^F)}{v\hat{a}r(r_{t+1})} \quad (31)$$

where g_t^F and μ_t^F represent the state-estimates returned by the Kalman-Filtering process. The difference in the t -indices between the filtered values and the actual observations is explained by the fact that μ_t and g_t describe the expected values of returns and dividend-growth in the next period $t + 1$:

$$g_t = E[r_{t+1}] \quad \text{and} \quad \mu_t = E[\Delta d_{t+1}]$$

Note that \hat{g}_t is a direct product of the Kalman Filter recursion, while $\hat{\mu}_t$ is an implied value based on the present-value equation 12.

$$\begin{aligned} pd_t &= A - B_1\hat{\mu}_t + B_2\hat{g}_t \\ \Leftrightarrow \hat{\mu}_t &= \frac{A + B_2\hat{g}_t - pd_t}{B_1} \end{aligned}$$

We also need to consider that the Filter estimates the demeaned version of the expected dividend growth. Thus, to obtain the true estimates of r_{t+1} and Δd_{t+1} , we need to add the unconditional mean in form of γ_0 and γ_1

$$g_t = \hat{g}_t + \gamma_0 \quad \text{and} \quad \mu_t = \hat{\mu}_t + \delta_0$$

In this way, we obtain a series of values which we can subsequently use for the R^2 computation (equation 30 and 31). For the CDAX data set, I obtain the values reported in table 5. We obtain a surprisingly high R^2 value of 35.8 % for dividend growth-rates, while the goodness of fit for returns is quite poor with 2.7% in the annual time series. For the CRSP data set, I obtain higher R^2 values for dividend

growth compared to returns as well (11.7% vs. 9.0%). However, the difference is less significant, and the fit for returns is better. For the semi-annual series, both measures report lower values with 21.6% and 1.0% respectively. This is consistent with the general assumption that returns are less predictable on the short-term (Cochrane, 2009a).

Table 5: In-Sample R^2 Values (CDAX, 1990–2018)

	Annual Data	Semi-Annual Data
R_{Div}^2	35.8%	21.6%
R_{Ret}^2	2.7%	1.0%

Table 6 and table 7 show the results of simple predictive regressions. We can see that the present-value state-space model outperforms the simple regressions when it comes to dividend-growth estimation. However, in terms of return predictability, the present-value model seems to have no advantage compared to these basic tools.

Table 6: OLS Predictive Regressions (CDAX 1990-2018)

The table reports the OLS regression results of log returns and log dividend growth rates on the lagged log price-dividend ratio using data between 1990 and 2018. The first two columns report the results for the annual data set and the last two columns for the semi-annual data set. Two asterisks (**) indicate significance at the 5% level, and three asterisks indicate significance at the 1% level

Dependent Variable	Annual Data		Semi-Annual Data	
	r_t	Δd_t	r_t	Δd_t
Constant	0.6030	-0.9159**	0.13697	-0.56846**
pd_{t-1}	-0.1430	0.2558**	-0.02897	0.15911***
R^2	3.93%	28.94%	0.36%	18.98%
Adj R^2	0.24%	26.21%	-1.45%	17.51%

6.3 Hypothesis Testing

Provided with our optimal estimation parameters, I would like to assess the results by the application of appropriate hypothesis tests. Having obtained our optimal es-

Table 7: AR(1)-Regressions (CDAX 1990-2018)

The table reports the AR(1)-OLS regression results of log returns and log dividend growth rates using data between 1990 and 2018. The first two columns report the results for the annual data set and the last two columns for the semi-annual data set. Two asterisks (**) indicate significance at the 5% level, and three asterisks indicate significance at the 1% level

Dependent Variable	Annual Data		Semi-Annual Data	
	r_t	Δd_t	r_t	Δd_t
Constant	0.07352	0.04449	0.0247	0.0254
AR(1)- Component	-0.10473	-0.01175	0.11112	0.0901
R^2	1.08%	0.02%	1.20%	0.84%
Adj R^2	-2.72%	-3.83%	-0.60%	-0.96%

timates via the maximum-likelihood estimation, we can make use of the likelihood-ratio test. By using this ratio, I test the return predictability, dividend growth predictability and persistence of expected dividend growth and returns. The respective test statistic is given by

$$LR = 2(\mathcal{L}^1 - \mathcal{L}^0),$$

where \mathcal{L}^1 represents the unconstrained model and \mathcal{L}^0 the constrained one. The test statistic asymptotically follows a χ^2 -square distribution with as many degrees of freedom as there are constrained parameters. For each conducted test, the results of the constrained model estimation and the likelihood-ratio are presented in Appendix E.

First, I test for the existence of return predictability. If the return predictability would be absent within the model, expected returns should show no persistence nor variation or correlation towards other variables. Hence, the corresponding hypothesis

$$H_0 : \delta_1 = \sigma_\mu = \rho_{\mu g} = \rho_{\mu d} = 0$$

$$H_1 : \delta_1 \neq 0 \text{ or } \sigma_\mu \neq 0 \text{ or } \rho_{\mu g} \neq 0 \text{ or } \rho_{\mu d} \neq 0$$

The likelihood-ratio, which follows a χ_4^2 -distribution, returns values of 54.65 for the annual and 37.03 for the semi-annual CDAX data. Hence, we clearly reject the null-hypothesis. However, the discussed low R^2 values for the return series cause doubt on the level of existent predictability.

Second, I test for a lack of dividend growth predictability. Again, if dividend-growth predictability is absent, expected growth rates should have no persistence and no variation or correlation towards other variables, where ρ_{gd} is already set to zero, because of the identity problem discussed in section 5.5. Consequentially, the tested hypothesis is given by

$$H_0 : \gamma_1 = \sigma_g = \rho_{\mu g} = 0$$

$$H_1 : \gamma_1 \neq 0 \text{ or } \sigma_g \neq 0 \text{ or } \rho_{\mu g} \neq 0$$

The likelihood-ratio follows a χ_3^2 distribution in this case. Again, we clearly reject the hypothesis on a 1% level, which is further supported by a high R^2 value for the dividend growth.

Furthermore, I investigate whether the expected dividend growth is not persistent. This is equivalent to testing the hypothesis

$$H_0 : \gamma_1 = 0$$

$$H_1 : \gamma_1 \neq 0$$

The ratio follows a χ_1^2 distribution and we strongly reject the H_0 on a 1% significance level for both CDAX time series, as I expected because of the large γ_1 estimate.

Lastly, I test for the common assumption, that expected dividend-growth rates and expected returns are equally persistent (see Cochrane (2007); Pástor and Stambaugh (2009)). The corresponding hypothesis is

$$H_0 : \gamma_1 = \delta_1$$

$$H_1 : \gamma_1 \neq \delta_1$$

The ratio follows a χ_1^2 distribution and we reject the hypothesis. Again, the results is rather unsurprising considering the immense difference between the estimates of the two parameters.

6.4 Validity of the Results

The present-value model contains a few assumptions and definitions that have been criticized and further examined in the literature. In this section, I discuss the robustness of the model and the validity of my results.

First, the considered sample size is very small, especially for the annual data. Looking at the goodness of fit, the artificially increased sample size in the form of semi-annual observations does not lead to more robust results. Considering that Van Binsbergen and Kojen (2010) reports large standard-errors for the correlation-parameter $\rho_{\mu d}$ and $\rho_{\mu g}$, the error for my estimates could be consequently larger. A longer observation period might have led to deeper insights.

Second, the assumptions of the proposed state-space present-value model can be criticized. In the present-value model derivation, I made use of an approximation to obtain a log-linearized return relation. Consequentially, this could lead to an approximation error in the final model. A possible solution would be the application of the unscented Kalman Filter, which is not reliant on the linearity of the underlying system. Van Binsbergen and Kojen (2011) apply this concept on a non-linear state-space model. For the CRSP data set, they could not find a significant impact of an approximation error caused by the log-linearized version. However, it would be interesting to see if this also holds for my CDAX data set.

Another aspect which should be critically reviewed is the assumption of a constant risk in the Kalman Filter procedure. We assume that the covariance matrix of our model innovation Σ does not vary over time. This is unrealistic considering that the

volatility on the stock market is varying over time. Ang and Liu (2007) show in their paper that the price-dividend ratio and expected returns are heteroscedastic. Piatti and Trojani (2017) adapts this thought and creates a state-space representation of the present-value model, that contains a time-varying risk component. They find differences in the persistence of dividend-growth rates compared to the homoscedastic model. A time-varying risk component could eventually also help to understand the negative correlation between expected returns and expected dividend growth in our model. Therefore, it would be interesting to apply the model of Piatti and Trojani on the CDAX data set.

7 Conclusion

In the thesis, I replicate and derive the state-space representation of the present-value model according to Van Binsbergen and Kojien (2010). Because of the limited available sample size for yearly observations of the CDAX, I construct a semi-annual time series of the CDAX and compare the results. I further construct the Simulated Annealing algorithm to solve the Maximum Likelihood optimization problem and obtain optimal estimates based on the CDAX data set.

I find a high level of persistence in the expected stock returns and lower levels for expected dividend growth rates. This is consistent with the findings of Van Binsbergen and Kojien on the American stock market. On the other hand, we surprisingly find a strong negative correlation between expected dividend growth and expected returns in both of the time series. This speaks against the intuition of the present-value identity, which expects an increase in returns to come along with higher dividends. It also opposes the estimates which is found in the CRSP data set. It would be interesting to examine whether this negative correlation remains when the model is applied on a longer annual time series of the German stock market.

The in-sample performance shows a two-sided result. The model seems to estimate realized dividend growth rates efficiently, but lacks quality in terms of the return estimates. In the application into the CRSP data set, the difference in the the goodness of fit is similar for returns and growth rates (R^2 -value of 9% and 12% respectively). However, with regards to the CDAX, we obtain a R^2 value of 35.8% for dividend growth and a value of only 2.7% for returns, which questions the model's ability to estimate returns. On the other hand, the benchmark models in the form of simple predictive regressions perform worse in the estimation of dividend growths while they also fail to model returns appropriately.

Last but not least, I conduct several hypothesis tests to detect predictability of returns and dividend growth rates. The likelihood-ratio tests strongly reject the hy-

potheses that there is no return predictability nor dividend growth predictability in the data, which is consistent with the findings in the CRSP data set. The tests further underline the persistence of dividend growth.

Further research via the state-space representation of the present-value model should try to apply the method to a larger sample to detect long-term trends in the data. Considering the heteroscedasticity in the stock data, an application of model with a time varying risk component would be interesting to examine. Furthermore, the restriction on a linear system through the requirements of the Kalman Filter might weaken the model's performance. Therefore, another potential extension would include the application of the unscented Kalman Filter to analyze non-linear relationships in the data.

Bibliography

- Ang, A., and G. Bekaert. 2006. Stock return predictability: Is it there? *The Review of Financial Studies* 20:651–707.
- Ang, A., and J. Liu. 2004. How to discount cashflows with time-varying expected returns. *Journal of Finance* 59:2745–2783.
- Ang, A., and J. Liu. 2007. Risk, return, and dividends. *Journal of Financial Economics* 85:1–38.
- Asimakopoulos, P., S. Asimakopoulos, N. Kourogenis, and E. Tsiritakis. 2017. Time-Disaggregated Dividend–Price Ratio and Dividend Growth Predictability in Large Equity Markets. *Journal of Financial and Quantitative Analysis* 52:2305–2326.
- Ball, R. 1978. Anomalies in relationships between securities' yields and yield-surrogates. *Journal of financial economics* 6:103–126.
- Bansal, R., and A. Yaron. 2004. Risks for the long run: a potential resolution of asset pricing puzzles. *Journal of Finance* 59:1481–1509.
- Bekaert, E., Geert. Engstron, and S. R. Grenadier. 2001. Stock and bond pricing in an affine economy. *NBER Working paper* 7346.
- Bollerslev, T., J. Marrone, L. Xu, and H. Zhou. 2014. Stock return predictability and variance risk premia: statistical inference and international evidence. *Journal of Financial and Quantitative Analysis* 49:633–661.
- Brandt, M. W., and Q. Kang. 2004. On the relation between the conditional mean and volatility of stock returns: A latent VAR approach. *Journal of Financial Economics* 72:217–257.

- Campbell, J., and J. Cochrane. 1999. By force of habit: A consumption-based explanation of aggregate stock market behavior. *Journal of Political Economy* 107:205–251.
- Campbell, J. Y., and R. J. Shiller. 1988a. The dividend-price ratio and expectations of future dividends and discount factors. *The Review of Financial Studies* 1:195–228.
- Campbell, J. Y., and R. J. Shiller. 1988b. Stock prices, earnings, and expected dividends. *The Journal of Finance* 43:661–676.
- Chen, L. 2009. On the reversal of return and dividend predictability. *Journal of Financial Economics* 92:128–151.
- Cochrane, J. H. 1992. Explaining the variance of price–dividend ratios. *The Review of Financial Studies* 5:243–280.
- Cochrane, J. H. 2007. The dog that did not bark: A defense of return predictability. *The Review of Financial Studies* 21:1533–1575.
- Cochrane, J. H. 2009a. *Asset pricing: Revised edition*. Princeton university press.
- Cochrane, J. H. 2009b. *Asset pricing: Revised edition*. Princeton university press.
- Cochrane, J. H. 2011. Presidential address: Discount rates. *The Journal of finance* 66:1047–1108.
- Corana, A., M. Marchesi, C. Martini, and S. Ridella. 1987. Minimizing multimodal functions of continuous variables with the simulated annealing algorithm. *ACM Transactions on Mathematical Software (TOMS)* 13:262–280.
- De Cesari, A., and W. Huang-Meier. 2015. Dividend changes and stock price informativeness. *Journal of Corporate Finance* 35:1–17.
- Durbin, J., and S. J. Koopman. 2012. *Time series analysis by state space methods*. Oxford university press.

- Fama, E. 1991. Efficient Capital Markets: II. *Journal of Finance* 46:1575–1618.
- Fama, E. F. 1965. The Behavior of Stock-Market Prices. *The Journal of Business* 38:34–105.
- Fama, E. F., and K. R. French. 1988. Dividend yields and expected stock returns. *Journal of Financial Economics* 22:3–25.
- Gordon, M. J. 1959. Dividends, earnings, and stock prices. *The review of economics and statistics* 41:99–105.
- Goyal, A., and I. Welch. 2003. Predicting the equity premium with dividend ratios. *Management Science* 49:639–654.
- Hamilton, J. 1994. *Time series analysis*. Princeton: Princeton University Press.
- Harvey, A. C. 1990. *Forecasting, structural time series models and the Kalman filter*. Cambridge university press.
- Kalman, R. E. 1960. A new approach to linear filtering and prediction problems. *Journal of basic Engineering* 82:35–45.
- Koijen, R. S., and S. Van Nieuwerburgh. 2011. Predictability of returns and cash flows. *Annu. Rev. Financ. Econ.* 3:467–491.
- Lettau, M., and S. Van Nieuwerburgh. 2007. Reconciling the return predictability evidence: The review of financial studies: Reconciling the return predictability evidence. *The Review of Financial Studies* 21:1607–1652.
- Lewellen, J. 2004. Predicting returns with financial ratios. *Journal of Financial Economics* 74:209–235.
- Lior Menzly, P. V., Tano Santos. 2004. Understanding Predictability. *Journal of Political Economy* 112:1–47.

- Lund, J., and T. Engsted. 1996. GMM and present value tests of the C-CAPM: evidence from the Danish, German, Swedish and UK stock markets. *Journal of International money and Finance* 15:497–521.
- Pástor, L., and R. F. Stambaugh. 2009. Predictive systems: Living with imperfect predictors. *Journal of Finance* 64:1583–1628.
- Pástor, L., and P. Veronesi. 2003. Stock valuation and learning about profitability. *Journal of Finance* 58:1749–1789.
- Pástor, L., and P. Veronesi. 2006. Was there a NASDAQ bubble in the late 1990s? *Journal of Financial Economics* 81:61–100.
- Piatti, I., and F. Trojani. 2017. Predictable risks and predictive regression in present-value models. *Saïd Business School WP* 11.
- Piatti, I., and F. Trojani. 2019. Dividend Growth Predictability and the Price–Dividend Ratio. *Management Science* .
- Rytchkov, O. 2012. Filtering out expected dividends and expected returns. *The Quarterly Journal of Finance* 2:1250012.
- Stambaugh, R. F. 1999. Predictive regressions. *Journal of Financial Economics* 54:375–421.
- Timmermann, A. 1993. How learning in financial markets generates excess volatility and predictability of stock returns. *The Quarterly Journal of Economics* 108:1135–1145.
- Van Binsbergen, J. H., and R. S. Koijen. 2010. Predictive regressions: A present-value approach. *The Journal of Finance* 65:1439–1471.
- Van Binsbergen, J. H., and R. S. Koijen. 2011. Likelihood-based estimation of exactly-solved present-value models. *Available at SSRN 1439849* .

- Wachter, J. A., and M. Warusawitharana. 2015. What is the chance that the equity premium varies over time? Evidence from regressions on the dividend-price ratio. *Journal of Econometrics* 186:74–93.
- Welch, I., and A. Goyal. 2008. A Comprehensive Look at The Empirical Performance of Equity Premium Prediction. *The Review of Financial Studies* 21:1455–1508.
- William L. Goffee, G. D. F., and J. Rogers. 1994. Global optimization of statistical functions with simulated annealing. *Journal of Econometrics* 60:65–99.
- Yongok Choi, S. J., and J. Y. Park. 2016. A reexamination of stock return predictability. *Journal of Econometrics* 192:168–189.

Appendices

A The Derivation of the Present-Value Model

The model used in this paper is based on the present-value identity by Campbell and Shiller (1988b). In the following, I am going to derive the log-linearized return relation of the price-dividend ratio and subsequently derive the present value equation, which is essential for my approach. We start by a simple return identity:

$$1 = R_{t+1}^{-1} * R_{t+1} = R_{t+1}^{-1} \cdot \frac{P_{t+1} + D_{t+1}}{P_t}$$

Multiplying by P_t/D_t results in

$$\begin{aligned} \frac{P_t}{D_t} &= R_{t+1}^{-1} \cdot \frac{P_{t+1} + D_{t+1}}{P_t} \cdot \frac{P_t}{D_t} = R_{t+1}^{-1} \cdot \frac{P_{t+1} + D_{t+1}}{D_t} \\ &= R_{t+1}^{-1} \cdot \left(\frac{P_{t+1}}{D_t} + \frac{D_{t+1}}{D_t} \right) = R_{t+1}^{-1} \cdot \left(1 + \frac{P_{t+1}}{D_t} \right) \cdot \frac{D_{t+1}}{D_t} \end{aligned}$$

Taking the logs will result in a log-linearized expression of the price-dividend ration pd_t :

$$\ln \left(\frac{P_t}{D_t} \right) = p_t - d_t = pd_t = \ln(R_{t+1}^{-1}) + \ln \left(1 + \frac{P_{t+1}}{D_t} \right) + \ln \left(\frac{D_{t+1}}{D_t} \right)$$

Using the property $P_t/D_t = \exp(\ln(P_t/D_t)) = \exp(pd_t)$ leads to

$$pd_t = -r_{t+1} + \Delta d_{t+1} + \ln(1 + \exp(pd_{t+1}))$$

The last term, $\ln(1 + \exp(pd_{t+1}))$, can be treated with a first-order Taylor Expansion around some point $\bar{pd} = E[pd_t]$ to get the following approximation:

$$pd_t \simeq -r_{t+1} + \Delta d_{t+1} + \ln(1 + \exp(p_{t+1} - d_{t+1})) \quad (\text{A1})$$

$$\simeq -r_{t+1} + \Delta d_{t+1} + \ln(1 + \exp(\bar{pd})) + \frac{\exp(\bar{pd})}{1 + \exp(\bar{pd})} * [pd_t + 1 - \bar{pd}] \quad (\text{A2})$$

After substituting part of the terms for $\kappa = \ln(1 + \exp(\overline{pd})) - \rho\overline{pd}$ and $\rho = \frac{\exp(\overline{pd})}{\exp(1+\overline{pd})}$ the log-linearized price-dividend ratio relation is given by:

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

We can iterate this equation forward by gradually substituting for pd_{t+i} :

$$\begin{aligned} pd_t &= \kappa + \rho pd_{t+1} + \Delta d_{t+1} - r_{t+1} \\ &= \kappa + \rho(\kappa + \rho pd_{t+2} + \Delta d_{t+2} - r_{t+2}) + \Delta d_{t+1} - r_{t+1} \\ &= \kappa + \rho\kappa + \rho^2 pd_{t+2} + (\Delta d_{t+1} - r_{t+1}) + \rho(\Delta d_{t+2} - r_{t+2}) \\ &= \dots \\ &= \sum_{j=0}^{\infty} \rho^j \kappa + \rho^{\infty} pd_{\infty} + \sum_{j=1}^{\infty} \rho^{j-1} (\Delta d_{t+j} - r_{t+j}) \end{aligned}$$

Since $\rho < 1$ (Cochrane (2009b) reports a value of 0.96 based on a historical-average price-dividend ratio of 25), we can assume that

$$\rho^{\infty} pd_{\infty} = \lim_{j \rightarrow \infty} \rho^j pd_j = 0$$

and make use the properties of a infinite geometric series

$$\sum_{j=0}^{\infty} a^j b = \frac{b}{1-a} \quad (\text{A3})$$

Hence, we get:

$$\begin{aligned} pd_t &= \sum_{j=0}^{\infty} \rho^j \kappa + \rho^{\infty} pd_{\infty} + \sum_{j=1}^{\infty} \rho^{j-1} (\Delta d_{t+j} - r_{t+j}) \\ &= \frac{\kappa}{1-\rho} + \sum_{j=1}^{\infty} \rho^{j-1} (\Delta d_{t+j} - r_{t+j}) \end{aligned}$$

Now, we can take expectations conditional upon time to define a relation among the price-dividend ratio, expected returns and expected dividend growth. Because this

equation holds ex-ante and ex-post, the expectation operator can be added on the right-hand side:

$$\begin{aligned} pd_t &= E_t \left[\frac{\kappa}{1-\rho} + \sum_{j=1}^{\infty} \rho^{j-1} (\Delta d_{t+j} - r_{t+j}) \right] \\ &= \frac{\kappa}{1-\rho} + \sum_{j=1}^{\infty} \rho^{j-1} E_t [\Delta d_{t+j} - r_{t+j}] \end{aligned}$$

By inserting $\mu_t = E_t[r_{t+1}]$ and $g_t = E_t[\Delta d_{t+1}]$ and adjusting the index of the sum operator, we find:

$$\begin{aligned} pd_t &= \frac{\kappa}{1-\rho} + \sum_{j=1}^{\infty} \rho^{j-1} E_t [g_{t+j-1} - \mu_{t+j-1}] \\ &= \frac{\kappa}{1-\rho} + \sum_{j=0}^{\infty} \rho^j E_t [g_{t+j} - \mu_{t+j}] \end{aligned}$$

We can now make use of the AR(1)-properties of the expected returns and expected dividend growth defined in equation (10):

$$\begin{aligned} pd_t &= \frac{\kappa}{1-\rho} + \sum_{j=0}^{\infty} \rho^j E_t [g_{t+j} - \mu_{t+j}] \\ &= \frac{\kappa}{1-\rho} + \sum_{j=0}^{\infty} \rho^j [\gamma_0 + \gamma_1(g_{t+j-1} - \gamma_0) - \delta_0 - \delta_1(\mu_{t+j-1} - \delta_0)] \end{aligned}$$

Note that the error-terms of the AR-processes have zero mean and can be omitted after taking expectations. We can now substitute for g_t and μ_t and iterate the two terms forward considering the following property of the AR-process:

$$\begin{aligned} \mu_{t+1} &= \delta_0 + \delta_1(\mu_t - \delta_0) + \epsilon_{t+1}^\mu \\ \Leftrightarrow E[\mu_{t+1}] &= \delta_0 + \delta_1(\mu_t - \delta_0) \\ \Leftrightarrow E[\mu_{t+2}] &= \delta_0 + \delta_1(\mu_{t+1} - \delta_0) \\ &= \delta_0 + \delta_1[\delta_0 + \delta_1(\mu_t - \delta_0) - \delta_0] \\ &= \delta_0 + \delta_1^2(\mu_t - \delta_0) \\ \Leftrightarrow E[\mu_{t+3}] &= \delta_0 + \delta_1^3(\mu_t - \delta_0) \\ \Leftrightarrow E[\mu_{t+j}] &= \delta_0 + \delta_1^j(\mu_t - \delta_0) \end{aligned} \tag{A4}$$

The same can be applied to the AR-process of the expected dividend growth and so we reach

$$\begin{aligned}
 pd_t &= \frac{\kappa}{1-\rho} + \sum_{j=0}^{\infty} \rho^j [\gamma_0 + \gamma_1(g_{t+j-1} - \gamma_0) - \delta_0 - \delta_1(\mu_{t+j-1} - \delta_0)] \\
 &= \frac{\kappa}{1-\rho} + \sum_{j=0}^{\infty} \rho^j [\gamma_0 + \gamma_1^j(g_t - \gamma_0) - \delta_0 - \delta_1^j(\mu_t - \delta_0)] \\
 &= \frac{\kappa}{1-\rho} + \sum_{j=0}^{\infty} \rho^j(\gamma_0 - \delta_0) + \sum_{j=0}^{\infty} \rho^j [\gamma_1^j(g_t - \gamma_0) - \delta_1^j(\mu_t - \delta_0)]
 \end{aligned}$$

We can now make use of the properties of infinite geometric series:

$$\begin{aligned}
 pd_t &= \frac{\kappa}{1-\rho} + \frac{\gamma_0 - \delta_0}{1-\rho} + \sum_{j=0}^{\infty} \rho^j [\gamma_1^j(g_t - \gamma_0) - \delta_1^j(\mu_t - \delta_0)] \\
 &= \frac{\kappa}{1-\rho} + \frac{\gamma_0 - \delta_0}{1-\rho} + \sum_{j=0}^{\infty} \rho^j \gamma_1^j(g_t - \gamma_0) - \sum_{j=0}^{\infty} \rho^j \delta_1^j(\mu_t - \delta_0) \\
 &= \frac{\kappa}{1-\rho} + \frac{\gamma_0 - \delta_0}{1-\rho} + \frac{g_t - \gamma_0}{1-\rho\gamma_1} - \frac{\mu_t - \delta_0}{1-\rho\delta_1}.
 \end{aligned}$$

From here we can form the final present-value equation which connects the price-dividend ratio, the expected returns and expected dividend growth:

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

with

$$\begin{aligned}
 A &= \frac{\kappa}{1-\rho} + \frac{\gamma_0 - \delta_0}{1-\rho} \\
 B_1 &= \frac{1}{1-\rho\delta_1} \\
 B_2 &= \frac{1}{1-\rho\gamma_1}
 \end{aligned}$$

B The Kalman Filter

In time series analysis, filtering describes the process of treating data by removing unwanted components. In finance, one of the most applied filtering algorithms is the Kalman Filter, which is set up on the base of a state-space model. It proved to be very useful for noisy observations as we often find them in many economic time series systems. At each step in time, the Kalman filter is able to generate optimal estimates for the unobservable state parameters of a system (in our case the expected dividend growth). Further, it can compute predictions of the state variables. The generated estimates are recursively adjusted at each step in time based on the incoming observable measurements. In the following section, I will derive the Kalman filter with respect to the two defined present-value state-space representations. I will mainly stick to the elaborations described in Hamilton (1994) and Durbin and Koopman (2012), but will adjust the notation to the case presented in this thesis.

For transparency, we transform the state and the measurement equation of our state-space model in a general compact form as described in Hamilton (1994):

$$X_{t+1} = FX_t + v_{t+1}, \quad (\text{State Equation})$$

$$Y_t = C'z_t + M_2X_t, \quad (\text{Observation Equation})$$

where

$$\Gamma \epsilon_{t+1}^X = v_{t+1}, \quad \text{with } E[v_t v_\tau'] = \begin{cases} Q = \Gamma \Sigma \Gamma' & \text{for } t = \tau \\ 0 & \text{otherwise} \end{cases},$$

$$C'z_t = M_0 + M_1Y_{t-1},$$

$$\text{with } C' = \begin{bmatrix} \gamma_0 & 0 \\ (1 - \delta_1)A & \delta_1 \end{bmatrix}, \quad \text{and } z_t = \begin{bmatrix} 1 \\ pd_t \end{bmatrix}.$$

The dimensions of the used vectors and matrices can differ depending on the specifications of the present value model (such as the considered dividend-reinvestment

strategy). I will therefore derive the set-up for a general case in which X_t, Y_t and z_t are vectors of the dimension $(r \times 1)$, $(n \times 1)$ and $(k \times 1)$. Correspondingly, the innovations vector v_{t+1} is a $(r \times 1)$ vector as well. F, C' and M_2 represent predetermined matrices of the dimensions $(r \times r)$, $(n \times k)$ and $(n \times r)$. To start, I need to make a few assumptions, which are critical for further derivation. First, I assume that z_t has no information about the future values of X_{t+s} besides the one contained in the past observations of Y_t . I further assume that the error-vector v_t at time t contains no info about the initial value of the state X_1 :

$$E[v_t, X_1] = 0 \text{ for } t = 1, 2, \dots, T$$

Iterating backwards within the AR state equation and making use of this assumption result in

$$\begin{aligned} X_t &= FX_{t-1} + v_t = F(FX_{t-2} + v_{t-1}) + v_t = F^2X_{t-2} + Fv_{t-1} + v_t \\ &= \dots = v_t + Fv_{t-1} + F^2v_{t-2} + \dots + F^{t-2}v_2 + F^{t-1}X_1 \end{aligned}$$

which implies

$$E[v_t, X'_\tau] = 0 \text{ for } \tau = t - 1, t - 2, \dots, 1$$

Considering the measurement equation, we can similarly derive

$$E[v_t Y'_\tau] = 0$$

One of the primary purposes of the Kalman Filter is to estimate the unobserved state X_t while we only know about the observations Y_t . The filter creates least square forecasts of X_t to handle this task:

$$\hat{X}_{t+1|t} = E[X_{t+1} | \Upsilon_t]$$

with $\Upsilon_t = Y_t, Y_{t-1}, \dots, Y_1, z_t, z_{t-1}, \dots, z_1$. The mean squared error coming along with these forecasts is denoted as

$$P_{t+1|t} = E \left[(X_{t+1} - \hat{X}_{t+1|t})(X_{t+1} - \hat{X}_{t+1|t})' \right]$$

At each step in time, these state-forecasts are used to generate forecasts for the observation Y_t itself. The forecast $\hat{Y}_{t|t-1}$ is compared to the actual Y_t and the resulting forecasting error is then used to update the state-forecast $X_{t|t-1}$ to $X_{t|t}$. These steps are calculated recursively for each step in time until the end of the sample $t = T$ is reached. How this is happening in detail is described in the following subsections and the summary at the end of this chapter. Lastly, I first assume the initial values $X_{0|0}$ and $P_{0|0}$ for the recursion of the filtering process to be given, but will later explain, how to calculate these.

B.1 Forecasting Y_t

One of our main goals is to calculate forecasts of our unobserved state X_t . Considering that z_t provides no additional information about X_t (besides the one contained in Υ_{t-1}) we can define this forecast estimate based on Υ_{t-1} as following:

$$E[X_t|z_t, \Upsilon_{t-1}] = E[X_t|\Upsilon_{t-1}] = \hat{X}_{t|t-1}$$

The Kalman Filter is making use of the relation between Y_t and X_t to reach an optimal estimate and forecast of the state. Therefore, we are going to forecast Y_t first, which will provide us with the base for generating estimates for $X_{t+1|t}$ and $P_{t+1|t}$ in the end:

$$\hat{Y}_{t|t-1} = E[Y_t|z_t, \Upsilon_{t-1}]$$

Inserting the measurement equation gives us:

$$E[Y_t|z_t, X_t] = C'z_t + M_2X_t$$

We can now make use of the law of iterated projections (for two random variables X and Y : $E[X] = E[E[X|Y]]$), which gives us our forecasting equation:

$$\begin{aligned}\hat{Y}_{t|t-1} &= C'z_t + M_2E[X_t|z_t, \Upsilon_{t-1}] \\ &= C'z_t + M_2\hat{X}_{t|t-1}\end{aligned}$$

The corresponding forecasting error is:

$$\begin{aligned}Y_t - \hat{Y}_{t|t-1} &= C'z_t + M_2X_t - C'z_t - M_2\hat{X}_{t|t-1} \\ &= M_2(X_t - \hat{X}_{t|t-1})\end{aligned}$$

and its mean squared error (MSE) is given by

$$\begin{aligned}E[(Y_t - \hat{Y}_{t|t-1})(Y_t - \hat{Y}_{t|t-1})'] &= E[M_2(X_t - \hat{X}_{t|t-1})(X_t - \hat{X}_{t|t-1})'M_2'] \\ &= M_2E[(X_t - \hat{X}_{t|t-1})(X_t - \hat{X}_{t|t-1})']M_2' \\ &= M_2P_{t|t-1}M_2'\end{aligned}$$

B.2 Updating the State

We still need to define how to forecast the state and how to subsequently update it in consideration of the prediction errors of the measurements. I will first explain how to do the latter and then continue with the generation of forecasts for the state. Ultimately, I will sum up the whole recursion in the final subsection.

We want to update and adjust the initially forecasted estimate of the state $X_{t|t-1}$ by evaluating the forecasting error of the measurement Y_t . This adjusted state is defined as:

$$\hat{X}_{t|t} = \hat{E}[X_t|y_t, z_t, \Upsilon_{t-1}] = \hat{E}[X_t|\Upsilon_t]$$

We can obtain the optimal estimate $\hat{X}_{t|t}$ by making use of the formulas for updating a linear projection via a block triangular factorization. The formula for three random vectors Y_1, Y_2 and Y_3 is provided below:

$$E[Y_3|Y_2, Y_1] = E[Y_3|Y_1] + H_{32}H_{22}^{-1} [Y_2 - \hat{E}[Y_2|Y_1]]$$

with

$$H_{22} = E[(Y_2 - E[Y_2|Y_1])[Y_2 - E[Y_2|Y_1]]']$$

$$H_{32} = E[(Y_3 - E[Y_3|Y_1])[Y_2 - E[Y_2|Y_1]]']$$

Using this scheme we obtain:

$$\begin{aligned} E[X_t|Y_t, z_t, \Upsilon_{t-1}] &= E[X_t|z_t, \Upsilon_{t-1}] + \\ &E[(X_t - E[X_t|z_t, \Upsilon_{t-1}])(Y_t - E[Y_t|z_t, \Upsilon_{t-1}])'] \cdot \\ &E[(Y_t - E[Y_t|z_t, \Upsilon_{t-1}])(Y_t - E[Y_t|z_t, \Upsilon_{t-1}])']^{-1} \cdot \\ &[Y_t - E[Y_t|z_t, \Upsilon_{t-1}]], \end{aligned}$$

which is equal to

$$\begin{aligned} \hat{X}_{t|t} &= \hat{X}_{t|t-1} + E[(X_t - \hat{X}_{t|t-1})(Y_t - \hat{Y}_{t|t-1})'] \\ &\cdot E[(Y_t - \hat{Y}_{t|t-1})(Y_t - \hat{Y}_{t|t-1})']^{-1} [Y_t - \hat{Y}_{t|t-1}] \end{aligned}$$

The second term can be expressed by the definitions of $P_{t|t-1}$, M_2 and the forecast error of Y_t :

$$\begin{aligned} E[(X_t - \hat{X}_{t|t-1})(Y_t - \hat{Y}_{t|t-1})'] &= E[(X_t - \hat{X}_{t|t-1})(M_2(X_t - \hat{X}_{t|t-1}))'] \\ &= E[(X_t - \hat{X}_{t|t-1})(X_t - \hat{X}_{t|t-1})'] M_2' = P_{t|t-1} M_2' \end{aligned}$$

Inserting this result and the formula of the forecasting error of Y_t leads to our final update equation for the state vector:

$$\hat{X}_{t|t} = \hat{X}_{t|t-1} + P_{t|t-1}M_2'[M_2P_{t|t-1}M_2']^{-1}[Y_t - C'z_t - M_2\hat{X}_{t|t-1}]$$

The corresponding MSE follows according to the rules of a linear projection:

$$E[(Y_3 - Y_3|Y_2, Y_1)(Y_3 - Y_3|Y_2, Y_1)'] = H_{33} - H_{32}H_{22}^{-1}H_{23}$$

with H_{32} and H_{22} as previously defined and

$$H_{33} = E[(Y_3 - Y_3|Y_1)(Y_3 - Y_3|Y_1)']$$

Using this concept we obtain

$$\begin{aligned} P_{t|t} &= E[(X_t - \hat{X}_{t|t})(X_t - \hat{X}_{t|t})'] \\ &= E[(X_t - \hat{X}_{t|t-1})(X_t - \hat{X}_{t|t-1})'] - \\ &\quad E[(X_t - \hat{X}_{t|t-1})(Y_t - \hat{Y}_{t|t-1})'] [E[(Y_t - \hat{Y}_{t|t-1})(Y_t - \hat{Y}_{t|t-1})']^{-1}] \cdot \\ &\quad E[(Y_t - \hat{Y}_{t|t-1})(X_t - \hat{X}_{t|t-1})'] \\ &= P_{t|t-1} - P_{t|t-1}M_2'(M_2P_{t|t-1}M_2')^{-1}M_2P_{t|t-1} \end{aligned}$$

B.3 Forecasting the State

Based on the obtained measures, we can now forecast the state vector for the next time step. Via the definition of the state equation we derive

$$\begin{aligned} \hat{X}_{t+1|t} &= E[X_{t+1}|\Upsilon_t] \\ &= E[FX_t + v_{t+1}|\Upsilon_t] \\ &= FE[X_t|Y_t] + E[v_{t+1}|\Upsilon_t] \\ &= F\hat{X}_{t|t} \end{aligned}$$

We can now insert the previously derived update equation for the state:

$$\begin{aligned}\hat{X}_{t+1|t} &= F \left[\hat{X}_{t|t-1} + P_{t|t-1} M_2' [M_2 P_{t|t-1} M_2']^{-1} [Y_t - C' z_t - M_2 \hat{X}_{t|t-1}] \right] \\ &= F \hat{X}_{t|t-1} + F P_{t|t-1} M_2' [M_2 P_{t|t-1} M_2']^{-1} [Y_t - C' z_t - M_2 \hat{X}_{t|t-1}]\end{aligned}$$

The term $P_{t|t-1} M_2' [M_2 P_{t|t-1} M_2']^{-1}$ is often named as Kalman Gain matrix in the filtering process and is denoted by K_t . The Kalman Gain tells us how much we should change our estimate according to the incoming measurement at time t . So we can write:

$$\hat{X}_{t+1|t} = F \hat{X}_{t|t-1} + F K_t [Y_t - C' z_t - M_2 \hat{X}_{t|t-1}]$$

Lastly, we can derive the corresponding MSE of the forecast. We make use of the previous finding $\hat{X}_{t+1|t} = F \hat{X}_{t|t}$ and the formulation of the state equation:

$$\begin{aligned}P_{t+1|t} &= E[(X_{t+1} - \hat{X}_{t+1|t})(X_{t+1} - \hat{X}_{t+1|t})'] \\ &= E[(F X_t + v_{t+1} - F \hat{X}_{t|t})(F X_t + v_{t+1} - F \hat{X}_{t|t})'] \\ &= E[(F(X_t - \hat{X}_{t|t}) + v_{t+1})(F(X_t - \hat{X}_{t|t}) + v_{t+1})']\end{aligned}$$

The covariance between X_t and v_{t+1} is again zero and so we can rewrite

$$P_{t+1|t} = F E[(X_t - \hat{X}_{t|t})(X_t - \hat{X}_{t|t})'] F' + E[v_{t+1} v_{t+1}']$$

and define

$$P_{t+1|t} = F P_{t|t} F' + Q$$

B.4 Initialization Values

Before we start the recursion of the filtering process, we need to define the starting values of the state and its covariance matrix for the initialization of the filter. These

values are not based on any observation and are simply given by the unconditional mean and the unconditional covariance matrix of the state:

$$X_{0|0} = E[X_0]$$

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

This is a common procedure in the literature even though some authors initialize the recursion with the forecasts $X_{1|0}$ and $P_{1|0}$ instead (as in (Hamilton, 1994)). While our definition starts the recursion by forecasting $X_{1|0}$ and $P_{1|0}$ based on $X_{0|0}$ and $P_{0|0}$, the ladder method starts by calculating the updated state $X_{1|1}$ and $P_{1|1}$ straight away. This can lead to slightly different results at the beginning of the filtering process. However, in both approaches, the unconditional mean and the associated MSE are the applied starting values. If the VAR(1)-process of the state equation is covariance-stationary, these initial values are relatively easy to obtain. A process is covariance-stationary if it has a constant mean and if the covariance between the two states at different points in time depends solely on the relative positions of the two terms:

1. $E[X_t] = \mu$ (independent of t)

2. $E[(X_t - \mu)(X_{t+k} - \mu)'] = \gamma(k)$ (solely dependent on k and not on t)

This condition is given if all eigenvalues of F are within the unit circle ($|\lambda| < 1$). Since we assume in our model that $\gamma_1 < 1$, the matrix F of our present-value model fulfills this criterion. The unconditional mean can then be derived by taking expectations in both sides of our state equation. With $E[v_t] = 0$, we obtain:

$$E[X_{t+1}] = E[FX_t + v_{t+1}] = FE[X_t]$$

Because of the constant mean of our process, we can rewrite these terms as

$$E[X_t] = FE[X_t]$$

$$\begin{aligned} \Leftrightarrow E[X_t] - FE[X_t] &= 0 \\ \Leftrightarrow (I_r - F)E[X_t] &= 0 \end{aligned}$$

Because we assume γ_1 to be less than one, F has no eigenvalue equal to one. This implies that the matrix resulting from $(I_r - F)$ is non-singular and that the only solution to the equation is the trivial one:

$$E[X_t] = 0$$

We can now obtain the unconditional variance of X_t by taking the expectation of the product $X_t X_t'$ and subsequently substituting with the equation for our state in the model. Considering $E[v_t X_t'] = 0$ and $E[v_t] = 0$ we derive

$$\begin{aligned} P_{0|0} &= E[(X_t - E[X_t])(X_t - E[X_t])'] \\ &= E[X_t X_t'] \\ &= E[(FX_t + v_{t+1})(FX_t + v_{t+1})'] \\ &= E[(FX_t + v_{t+1})(X_t' F' + v_{t+1}')'] \\ &= E[FX_t X_t' F' + v_{t+1} X_t' F' + FX_t v_{t+1}' + v_{t+1} v_{t+1}'] \\ &= E[FX_t X_t' F'] + E[v_{t+1} v_{t+1}'] \end{aligned}$$

After defining $G = E[X_t X_t']$ (a constant term because of the stationarity) we can rewrite these terms and obtain the equation:

$$G = FGF' + G$$

which can be solved via the application of a vectorization and the Kronecker Product (denoted by $X \otimes Y$, see Section C). Making use of the property $vec(ABC) = (C' \otimes A)vec(B)$ we get:

$$\begin{aligned} G &= FGF' + G \\ \Leftrightarrow vec(G) &= vec(FGF') + vec(G) \end{aligned}$$

$$\begin{aligned} &\Leftrightarrow \text{vec}(G) = (F \otimes F)\text{vec}(G) + \text{vec}(Q) \\ \Leftrightarrow \text{vec}(G) - (F \otimes F)\text{vec}(G) &= \text{vec}(Q) \\ \Leftrightarrow [I_{r^2} - (F \otimes F)]\text{vec}(G) &= \text{vec}(Q) \\ \Leftrightarrow \text{vec}(G) &= [I_{r^2} - (F \otimes F)]^{-1}\text{vec}(Q) \end{aligned}$$

This represents the unconditional covariance matrix of our starting value and we can define:

$$\text{vec}(P_{0|0}) = [I_{r^2} - (F \otimes F)]^{-1}\text{vec}(Q)$$

B.5 Summary of the Kalman Filter for the Present-Value Model

We can now sum up the filtering process of our Present-Value State-Space model.

The initial state vector is defined as

$$X_{0|0} = E[X_1] = E[X_t] = 0_{rx1}$$

with the corresponding MSE / Covariance matrix

$$P_{0|0} = E[(X_t - E[X_t])(X_t - E[X_t])'] = E[X_t X_t']$$

which can be solved via the formula

$$\text{vec}(P_{0|0}) = [I_{r^2} - (F \otimes F)]^{-1}\text{vec}(Q)$$

We then start the filtering loop by generating forecasts of the state and calculating the corresponding MSE:

$$\begin{aligned} \hat{X}_{t+1|t} &= F\hat{X}_{t|t-1} + FK_t[Y_t - C'z_t - M_2\hat{X}_{t|t-1}] \\ P_{t|t-1} &= FP_{t-1|t-1}F' + Q \end{aligned}$$

with the Kalman Gain matrix defined as

$$K_t = P_{t|t-1}M_2'[M_2P_{t|t-1}M_2']^{-1}$$

The forecasts for the observations are then given by

$$\hat{Y}_{t|t-1} = C'z_t + M_2\hat{X}_{t|t-1}$$

To keep the following formulas simple and stick to the implementation in the code, I notate the corresponding forecasting error and its associated MSE η_t and S_t . They are given by:

$$\begin{aligned}\eta_t &= Y_t - C'z_t - M_2\hat{X}_{t|t-1} \\ S_t &= E[(Y_t - \hat{Y}_{t|t-1})(Y_t - \hat{Y}_{t|t-1})'] = M_2P_{t|t-1}M_2'\end{aligned}$$

Lastly, the updated state and its MSE are calculated as following:

$$\begin{aligned}\hat{X}_{t|t} &= \hat{X}_{t|t-1} + P_{t|t-1}M_2'[M_2P_{t|t-1}M_2']^{-1}[Y_t - C'z_t - M_2\hat{X}_{t|t-1}] \\ &= \hat{X}_{t|t-1} + K_t\eta_t \\ P_{t|t} &= P_{t|t-1} - P_{t|t-1}M_2'(M_2P_{t|t-1}M_2')^{-1}M_2P_{t|t-1} \\ &= (I_r - K_tM_2)P_{t|t-1}\end{aligned}$$

If we re-substitute the terms

$$\Gamma\epsilon_{t+1}^X = v_{t+1}, \text{ with } E[v_t v_t'] = \begin{cases} Q = \Gamma\Sigma\Gamma' \text{ for } t = \tau \\ 0 \text{ otherwise} \end{cases}$$

$$C'z_t = M_0 + M_1Y_{t-1}$$

we end up with the same process as described in Van Binsbergen and Koijen (2010):

$$X_{0|0} = E[X_0] = 0_r$$

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

$$\begin{aligned}
X_{t|t-1} &= FX_{t-1|t-1} \\
P_{t|t-1} &= FP_{t-1|t-1}F' + \Gamma\Sigma\Gamma' \\
\eta_t &= Y_t - M_0 - M_1Y_{t-1} - M_2X_{t|t-1} \\
S_t &= M_2P_{t|t-1}M_2' \\
K_t &= P_{t|t-1}M_2'S_t^{-1} \\
X_{t|t} &= X_{t|t-1} + K_t\eta_t \\
P_{t|t} &= (I_r - K_tM_2)P_{t|t-1}
\end{aligned}$$

where r describes the size of the state vector.

B.6 Maximum Likelihood Estimation of the Model Parameters

We have derived the forecasts $\hat{X}_{t|t-1}$ and $\hat{Y}_{t|t-1}$ in the sense of linear projections. They therefore present optimal linear forecasts conditional on z_t and Υ_{t-1} in any case. Furthermore, the errors in our state-space model are assumed to be normally distributed, which makes it possible to make an even stronger statement. Under these circumstances, the forecasts of our Kalman Filter are optimal in the light of any function of (z_t, Υ_{t-1}) . It also implies that Y_t conditional on (z_t, Υ_{t-1}) is normally distributed with mean $\hat{Y}_{t+1|t}$ and variance S_t :

$$Y_t|z_t, \Upsilon_{t-1} \sim N(\hat{Y}_{t+1|t}, S_t)$$

The corresponding density function is then given by

$$\begin{aligned}
f_{Y_t|z_t, \Upsilon_{t-1}}(\Delta d_t, pd_t) &= \frac{\exp\left[-\frac{1}{2}(Y_t - \hat{Y}_{t+1|t})'S_t^{-1}(Y_t - \hat{Y}_{t+1|t})\right]}{\sqrt{(2\pi)^2|S_t|}} \\
&= \frac{\exp\left(-\frac{1}{2}\eta_t'S_t^{-1}\eta_t\right)}{\sqrt{(2\pi)^2|S_t|}}
\end{aligned}$$

We can now easily derive the log-likelihood function:

$$l_t = \ln(f_{Y_t|z_t, \Upsilon_{t-1}}(\Delta d_t, pd_t))$$

$$\begin{aligned}
&= \ln \left[\exp \left(-\frac{1}{2} \eta_t' S_t^{-1} \eta_t \right) \right] - \ln \left(\sqrt{(2\pi)^2 |S_t|} \right) \\
&= -\frac{1}{2} \eta_t' S_t^{-1} \eta_t - \frac{1}{2} \ln \left((2\pi)^2 |S_t| \right) \\
&= -\frac{1}{2} \left[\eta_t' S_t^{-1} \eta_t + \ln(|S_t|) \right] - \ln(2\pi)
\end{aligned}$$

We can omit the last term and the constant factor since they will not affect the optimal solution which maximizes the likelihood:

$$l_t = -\eta_t' S_t^{-1} \eta_t - \ln(|S_t|)$$

We want to maximize the likelihood over the whole series of T observations, which is why we aim for maximization of the aggregated likelihood function. Ultimately, this provides us with the following likelihood function, which will be optimized via the Simulated Annealing algorithm:

$$L = -\sum_{t=1}^T \ln(|S_t|) - \sum_{t=1}^T \eta_t' S_t^{-1} \eta_t$$

Estimation Restrictions. Before getting started with the optimization of the likelihood function, a few constraints regarding the values of the estimation parameters need to be imposed. First, the Kalman filter comes along with an identity issue if we do not place any restrictions for F , Q , C and M_2 . If the parameters of our state-space model are unidentified, there is more than one set of parameter values that could result in the same likelihood-values. Consequentially, we would not be able to find the optimal parameter set for our present-value model. This is why we predetermine the correlation between the error terms for realized dividend growth and expected dividend growth to be zero:

$$\rho_{dg} = 0$$

Thereby, we make sure that all the parameters in the covariance matrix of the error-terms are identified.

Second, we need to set upper and lower boundaries for the rest of the parameters to be estimated. These will make sure that the covariance matrices in the model stay positive definite and that the AR-process of the state variable is covariance stationary. We ensure stationarity by defining:

$$|\gamma_1| < 1 \text{ and } |\delta_1| < 1$$

In the case of the cash-reinvestment model, the covariance matrix of the shocks stays positive definite by constraining the standard deviations of the shocks:

$$\sigma_g, \sigma_\mu, \sigma_d > 0$$

The correlation parameters are bound between -1 and 1:

$$-1 < \rho_{g\mu}, \rho_{\mu d} < 1$$

C The Kronecker Product

For the derivation of the initial covariance matrix of the Kalman Filter, I made use of the Kronecker Product (see Appendix C). For a $m \times n$ matrix A and $p \times q$, it is given by

$$A \otimes B = \begin{bmatrix} a_{11}\mathbf{B} & \cdots & a_{1n}\mathbf{B} \\ \vdots & \ddots & \vdots \\ a_{m1}\mathbf{B} & \cdots & a_{mn}\mathbf{B} \end{bmatrix}, \quad (\text{C1})$$

More explicitly, the result is a $mp \times nq$ -matrix (which explains the dimension of the unity matrix in equation 29):

$$A \otimes B = \begin{bmatrix} a_{11}b_{11} & a_{11}b_{12} & \cdots & a_{11}b_{1q} & \cdots & \cdots & a_{1n}b_{11} & a_{1n}b_{12} & \cdots & a_{1n}b_{1q} \\ a_{11}b_{21} & a_{11}b_{22} & \cdots & a_{11}b_{2q} & \cdots & \cdots & a_{1n}b_{21} & a_{1n}b_{22} & \cdots & a_{1n}b_{2q} \\ \vdots & \vdots & \ddots & \vdots & & & \vdots & \vdots & \ddots & \vdots \\ a_{11}b_{p1} & a_{11}b_{p2} & \cdots & a_{11}b_{pq} & \cdots & \cdots & a_{1n}b_{p1} & a_{1n}b_{p2} & \cdots & a_{1n}b_{pq} \\ \vdots & \vdots & & \vdots & \ddots & & \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots & & \ddots & \vdots & \vdots & & \vdots \\ a_{m1}b_{11} & a_{m1}b_{12} & \cdots & a_{m1}b_{1q} & \cdots & \cdots & a_{mn}b_{11} & a_{mn}b_{12} & \cdots & a_{mn}b_{1q} \\ a_{m1}b_{21} & a_{m1}b_{22} & \cdots & a_{m1}b_{2q} & \cdots & \cdots & a_{mn}b_{21} & a_{mn}b_{22} & \cdots & a_{mn}b_{2q} \\ \vdots & \vdots & \ddots & \vdots & & & \vdots & \vdots & \ddots & \vdots \\ a_{m1}b_{p1} & a_{m1}b_{p2} & \cdots & a_{m1}b_{pq} & \cdots & \cdots & a_{mn}b_{p1} & a_{mn}b_{p2} & \cdots & a_{mn}b_{pq} \end{bmatrix}. \quad (\text{C2})$$

D Estimation Results for the CRSP Data Set

D.1 Results for the CRSP time series from 1946–2018

Table D1: Annual Summary Statistics in the case of Cash-reinvested Dividends (CRSP, 1946–2018)

	Δd_t	r_t	pd_t
Mean	0.0609	0.1000	3.5000
Median	0.0544	0.1325	3.4514
Standard Deviation	0.0695	0.1639	0.4252
Maximum	0.2602	0.4082	4.4571
Minimum	-0.1875	-0.4830	2.7307
No. Observations	73	73	73

Table D2: Maximum-Likelihood Parameter Estimates (CRSP, 1946–2018)

Maximum Likelihood Estimates	
Parameter	Estimate
γ_0	0.061
δ_0	0.087
γ_1	0.263
δ_1	0.922
σ_g	0.067
σ_μ	0.016
σ_d	0.001
$\rho_{g\mu}$	0.253
$\rho_{\mu d}$	-0.913
Log-Likelihood-Value	519.102
Implied Present-Value Model Parameters	
A	3.627
B_1	9.487
B_2	1.344
ρ	0.971
\bar{R}_{Div}^2	11.7
\bar{R}_{Ret}^2	9.0

D.2 Replication: Comparison of the Results with the Present-Value model by Van Binsbergen and Koijen (2010)

Table D3: Summary Statistics of Annual Data (CRSP Data, 1946–2007)

	Δd_t	r_t	pd_t	Δd_t (Paper)
Mean	0.0613	0.1064	3.4417	0.0611
Median	0.0538	0.1335	3.3847	0.0540
Standard Deviation	0.0622	0.1549	0.4323	0.0622
Maximum	0.2602	0.4082	4.4571	0.2616
Minimum	-0.0555	-0.3277	2.7307	-0.0579
No. Observations	62	62	62	62

Table D4: Maximum-Likelihood Parameter - Estimates (CRSP, 1946–2007)

Maximum Likelihood Estimates		
Parameter	Estimate	Estimate (Paper)
γ_0	0.062	0.062
δ_0	0.088	0.090
γ_1	0.357	0.354
δ_1	0.929	0.932
σ_g	0.058	0.058
σ_μ	0.016	0.016
σ_d	0.001	0.002
$\rho_{g\mu}$	0.387	0.417
$\rho_{\mu d}$	-0.888	-0.147
Log-Likelihood-Value	459.207	
Implied Present-Value Model Parameters		
A	3.624	3.571
B_1	10.043	10.334
B_2	1.529	1.523
ρ	0.969	0.969
R_{Div}^2	13.4%	13.9%
R_{Ret}^2	8.7%	8.2%

E Likelihood-Ratio Test Results

Table E1: Likelihood-Ratio Test 1

Likelihood-Ratio-Test 1		
$H_0 : \delta_1 = \sigma_\mu = \rho_{g\mu} = \rho_{\mu d} = 0$		
Parameter	Estimate (Annual Data)	Estimate (Semi-Annual Data)
γ_0	0.0287	0.0216
δ_0	0.0531	0.0461
γ_1	0.7315	0.8787
δ_1	0	0
σ_g	0.0719	0.0221
σ_μ	0	0
σ_d	0.1278	0.1044
$\rho_{g\mu}$	0	0
$\rho_{\mu d}$	0	0
Log-Likelihood-Value (H_0)	322.31	353.73
Log-Likelihood-Value (H_1)	349.63	372.25
Likelihood-Ratio	54.65	37.03
Critical Value of χ_4^2 , 5% Level	9.49	9.49
Critical Value of χ_4^2 , 1% Level	13.28	13.28

Table E2: Likelihood-Ratio Test 2

Likelihood-Ratio-Test 2		
$H_0 : \gamma_1 = \sigma_g = \rho_{g\mu} = 0$		
Parameter	Estimate (Annual Data)	Estimate (Semi-Annual Data)
γ_0	0.0439	0.0281
δ_0	0.0674	0.0594
γ_1	0	0
δ_1	0.8444	0.9658
σ_g	0	0
σ_μ	0.0454	0.0101
σ_d	0.1502	0.1113
$\rho_{g\mu}$	0	0
$\rho_{\mu d}$	0.5064	0.5741
Log-Likelihood-Value (H_0)	328.68	355.82
Log-Likelihood-Value (H_1)	349.63	372.25
Likelihood-Ratio	41.92	32.86
Critical Value of χ_3^2 , 5% Level	7.81	7.81
Critical Value of χ_3^2 , 1% Level	11.34	11.34

Table E3: Likelihood-Ratio Test 2

Likelihood-Ratio-Test 3		
$H_0 : \gamma_1 = 0$		
Parameter	Estimate (Annual Data)	Estimate (Semi-Annual Data)
γ_0	0.0404	0.0271
δ_0	0.0638	0.0521
γ_1	0	0
δ_1	0.7467	0.9105
σ_g	0.1504	0.0963
σ_μ	0.0505	0.0150
σ_d	0.0130	0.0558
$\rho_{g\mu}$	0.1422	0.1166
$\rho_{\mu d}$	-0.3626	0.3553
Log-Likelihood-Value (H_0)	342.39	366.43
Log-Likelihood-Value (H_1)	349.63	372.25
Likelihood-Ratio	14.49	11.63
Critical Value of χ_1^2 , 5% Level	3.84	3.84
Critical Value of χ_1^2 , 1% Level	6.63	6.63

Table E4: Likelihood-Ratio Test 2

Likelihood-Ratio-Test 4		
$H_0 : \gamma_1 = \delta_1$		
Parameter	Estimate (Annual Data)	Estimate (Semi-Annual Data)
γ_0	0.0366	0.0242
δ_0	0.0604	0.0480
γ_1	0.5958	0.8292
δ_1	0.5958	0.8292
σ_g	0.0692	0.0379
σ_μ	0.0448	0.0149
σ_d	0.1267	0.0967
$\rho_{g\mu}$	-0.6317	0.5743
$\rho_{\mu d}$	0.9302	0.9851
Log-Likelihood-Value (H_0)	339.95	366.87
Log-Likelihood-Value (H_1)	349.63	372.25
Likelihood-Ratio	19.38	10.76
Critical Value of χ_1^2 , 5% Level	3.84	3.84
Critical Value of χ_1^2 , 1% Level	6.63	6.63