Sequential Bayesian Learning in the Present Value Model

Artem Dyachenko

Central European University

Department of Economics

In partial fulfillment of requirements for the degree of Master of Arts

Supervisor: Professor Andras Fulop

Hungary, Budapest

2014

Contents

Abstract

In this thesis, I re-estimate the present value model with hidden variables of Binsbergen and Koijen (2010) using the sequential Monte Carlo algorithm instead of maximum likelihood procedure. The latent variables are expected returns and expected dividend growth. First, I show that in-sample forecasts are more optimistic than real-time, out-of-sample forecasts. Second, returns are close to being unpredictable out-of-sample, which implies that returns may follow a martingale process.

Acknowledgement

I thank Andras Fulop for his help in writing this thesis. I thank my mother for her support and pushing me to finish this thesis.

Chapter 1

Introduction

The present value model shows that price dividend ratio of the aggregate stock market is a linear function of expected future dividend growth rates and expected returns of the aggregate stock market. Therefore, the price-dividend ratio can be used to forecast both future dividends and returns. Binsbergen & Koijen (2010) develop an approach about how to use price-dividend ratio and dividend growth rate to predict jointly returns and dividends. They introduce unobserved variables into the present value model: expected returns and expected dividend growth rates which are assumed to follow AR(1) processes. To find parameters of their extended present value model, the authors define the transition and measurement equations of the model and run maximum likelihood. Given parameters of the model and the data (which is the series of price dividend ratio and dividend growth rate), it is possible to estimate the time series for hidden variables with the Kalman filter. In addition, Binsbergen & Koijen (2010) show that next period's dividend growth and return are the linear functions of the whole history of dividend growth rates and price dividend ratios, when the Kalman filter is used to uncover hidden variables. However, maximum likelihood treats parameters as unknown and fixed numbers.

The purpose of this thesis is to re-estimate the present value model with latent variables using sequential Monte Carlo estimator instead of maximum likelihood procedure to treat parameters as random variables and thus to account for parameter uncertainty. Sequential Monte Carlo estimation will also allow me to see how well the present value model with hidden variables forecasts in-sample and out-of-sample.

This work relates to the rest of the literature about predicting returns and dividend growth rates. A common observation is that returns are predictable using price dividend ratio while dividend growth rate is not. Since price-dividend ratio can be expressed as the linear function of future returns and future dividend growth rates and since price-dividend ratio has variability, then price-dividend should vary either due to future return or future dividend growth. Cochrane (2008) argues that returns are predictable because dividend growth rates are not by using different statistical tests. Binsbergen and Koijen (2010) point out that you can forecast both returns and dividend growth by using price-dividend ratio. However, Binsbergen and Koijen (2010) also note that probability density function of persistence coefficient for expected dividends is multimodal. In addition, the estimate of the persistence coefficient for expected dividend growth has relatively big standard errors.

In this thesis, I assume that steady state price-dividend value does not change over time. Lettau and Van Nieuwerburgh (2008) show that incorporating shifts in price-dividend steady state value is important and can improve an in-sample forecast. However, since the magnitude and time of the steady state shifts are uncertain, it's hard to use steady state dynamics to forecast in real time.

The thesis has the following structure. In Chapter 2 I re-derive the present value model with latent variables. In Chapter 3 I derive the Kalman filter and explain the sequential Monte Carlo (SMC) algorithm. In Chapters 4 and 5, I present the simulation results and real data estimation results respectively. Chapter 6 concludes.

Chapter 2

Present Value Model

2.1 Basic Present-Value Model

In this section I derive the present value model. My derivation follows closely an explanation of present value model in Cochrane (2005).

$$1 = R_{t+1}^{-1} R_{t+1} = R_{t+1}^{-1} \frac{P_{t+1} + D_{t+1}}{P_t} \tag{2.1.1}$$

$$\frac{P_t}{D_t} = [R_{t+1}^{-1} \frac{P_{t+1} + D_{t+1}}{P_t}] \frac{P_t}{D_t} = R_{t+1}^{-1} (\frac{P_{t+1} + D_{t+1}}{D_t}) = R_{t+1}^{-1} (1 + \frac{P_{t+1}}{D_{t+1}}) \frac{D_{t+1}}{D_t} \tag{2.1.2}$$

I write (2.1.2) in terms of logs to get

$$p_t - d_t = -r_{t+1} + \Delta d_{t+1} + \log \left[1 + \frac{P_{t+1}}{D_{t+1}} \right] \tag{2.1.3}$$

where

$$\Delta d_{t+1} \equiv \log \left[\frac{D_{t+1}}{D_t} \right]$$

and

$$r_{t+1} \equiv \log\left[\frac{P_{t+1} + D_{t+1}}{P_t}\right]$$

.

Using Taylor-series approximation around $p - d$, I can write the last term in (2.1.3) as

$$\begin{aligned}
\log\left[1 + \frac{P_{t+1}}{D_{t+1}}\right] &= \log(1 + e^{p_{t+1} - d_{t+1}}) \\
&\approx \log(1 + e^{p-d}) + \frac{e^{p-d}}{1 + e^{p-d}}\left[(p_{t+1} - d_{t+1}) - (p - d)\right] \\
&= \log(1 + \frac{P}{D}) + \frac{\frac{P}{D}}{1 + \frac{P}{D}}\left[(p_{t+1} - d_{t+1}) - (p - d)\right] \\
&= \log(1 + \frac{P}{D}) - \frac{\frac{P}{D}}{1 + \frac{P}{D}}(p - d) + \frac{\frac{P}{D}}{1 + \frac{P}{D}}(p_{t+1} - d_{t+1}) \\
&= k + \rho(p_{t+1} - d_{t+1})
\end{aligned} \tag{2.1.4}$$

where

$$k \equiv \log(1 + \frac{P}{D}) - \frac{\frac{P}{D}}{1 + \frac{P}{D}}(p - d)$$

and

$$\rho \equiv \frac{\frac{P}{D}}{1 + \frac{P}{D}}$$

Then I plug (2.1.4) into (2.1.3) to get

$$p_t - d_t = -r_{t+1} + \Delta d_{t+1} + \underbrace{k + \rho(p_{t+1} - d_{t+1})}_{\log\left[1+\frac{P_{t+1}}{D_{t+1}}\right]} \tag{2.1.5}$$

I can rewrite (2.1.5) as

$$(p_t - d_t) - \rho(p_{t+1} - d_{t+1}) = -r_{t+1} + \Delta d_{t+1} + k \tag{2.1.6}$$

Let F denote a forward operator. Then I can rewrite (2.1.6) as

$$(1 - \rho F)(p_t - d_t) = -r_{t+1} + \Delta d_{t+1} + k \tag{2.1.7}$$

Assuming that $0 < \rho < 1$, then $(1 - \rho F)$ is an invertible polynomial and I can write equation (2.1.7) in the following way:

$$(p_t - d_t) = (1 - \rho F)^{-1}(-r_{t+1} + \Delta d_{t+1} + k) \tag{2.1.8}$$

The equation (2.1.8) is equivalent to

$$\begin{aligned}
(p_t - d_t) &= (1 + \rho F + \rho^2 F + ...)(-r_{t+1} + \Delta d_{t+1} + k) \\
&= (1 + \rho F + \rho^2 F + ...)\left[k + (\Delta d_{t+1} - r_{t+1})\right] \\
&= \frac{k}{1-\rho} + \sum_{j=1}^{\infty} \rho^{j-1}(\Delta d_{t+j} - r_{t+j})
\end{aligned} \tag{2.1.9}$$

The expression (2.1.9) says that pd_t is a a linear function of expected future dividends and returns.

2.2 Introducing expected return and expected dividend growth

This section introduces hidden variables as in the article by Binsbergen and Koijen (2010): expected return and expected dividend growth rate. While re-deriving, I borrow the notation from Binsbergen and Koijen (2010).

They define expected log-return $\mu_t$as

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

and expected log-dividend growth rate $g_t$as

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

where

$$\Delta d_{t+1} = g_t + \epsilon^d_{t+1} \tag{2.2.1}$$

.

Let expected returns (μ_t) and expected dividend growth rates (g_t) follow an AR(1) processes :

$$(\mu_{t+1} - \delta_0) = \delta_1(\mu_t - \delta_0) + \epsilon^{\mu}_{t+1} \tag{2.2.2}$$

$$(g_{t+1} - \gamma_0) = \gamma_1(g_t - \gamma_0) + \epsilon^{g}_{t+1} \tag{2.2.3}$$

where $\delta_0$and γ_0 are the means of expected return and expected dividend growth rate respectively.

If I forecast μ_{t+j}, then I set future shocks equal to zero and I get the following:

$$\begin{aligned}
(\mu_{t+1} - \delta_0) &= \delta_1(\mu_t - \delta_0) \\
(\mu_{t+2} - \delta_0) &= \delta_1(\mu_{t+1} - \delta_0) = \delta_1^2(\mu_t - \delta_0) \\
\vdots &\ \vdots\ \vdots \\
(\mu_{t+j} - \delta_0) &= \delta_1^j(\mu_t - \delta_0)
\end{aligned}$$

Similarly, if I forecast g_{t+j}, then I set future shocks equal to zero and I get the following:

$$\begin{aligned}
(g_{t+1} - \gamma_0) &= \gamma_1(g_t - \gamma_0) \\
(g_{t+2} - \gamma_0) &= \gamma_1(g_{t+1} - \gamma_0) = \gamma_1^2(g_t - \gamma_0) \\
\vdots &\ \vdots\ \vdots \\
(g_{t+j} - \gamma_0) &= \gamma_1^j(g_t - \gamma_0)
\end{aligned}$$

Then I take expectation of (2.1.9) conditional upon information at time t:

$$\begin{aligned}
(p_t - d_t) &= \frac{k}{1-\rho} + \sum_{j=1}^{\infty} E_t[\rho^{j-1}(\Delta d_{t+j} - r_{t+j})] \\
&= \frac{k}{1-\rho} + \sum_{j=1}^{\infty}[\rho^{j-1}(g_{t+j-1} - \mu_{t+j-1})] \\
&= \frac{k}{1-\rho} + \sum_{j=0}^{\infty}[\rho^{j}(g_{t+j} - \mu_{t+j})] \\
&= \frac{k}{1-\rho} + \sum_{j=0}^{\infty}\rho^{j}\left[\gamma_0 + \gamma_1^j(g_t - \gamma_0) - \delta_0 - \delta_1^j(\mu_t - \delta_0)\right] \\
&= \frac{k}{1-\rho} + \sum_{j=0}^{\infty}\rho^{j}\left[(\gamma_0 - \delta_0) - \delta_1^j(\mu_t - \delta_0) + \gamma_1^j(g_t - \gamma_0)\right] \\
&= \frac{k}{1-\rho} + \frac{\gamma_0 - \delta_0}{1-\rho} - \frac{\mu_t - \delta_0}{1-\rho\delta_1} + \frac{g_t - \gamma_0}{1-\rho\gamma_1} \\
&= A - B_1(\mu_t - \delta_0) + B_2(g_t - \gamma_0) \qquad (2.2.4)
\end{aligned}$$

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

The expression above says that current price-dividend ratio is a linear function of expected return and expected dividend growth for next period.

The shocks ϵ^d_{t+1}, ϵ^g_{t+1}, ϵ^μ_{t+1} from equations (2.1.1), (2.1.3) and (2.1.2) respectively are normally distributed, i.i.d. over time, with mean zero, and uncorrelated with each other. Also I assume that shocks are jointly normally distributed as in Binsbergen and Koijen (2010). The covariance matrix is given by

$$\Sigma = Var\left\{ \begin{bmatrix} \epsilon^d_{t+1} \\ \epsilon^g_{t+1} \\ \epsilon^\mu_{t+1} \end{bmatrix} \right\} = \begin{bmatrix} \sigma^2_\mu & 0 & 0 \\ 0 & \sigma^2_g & 0 \\ 0 & 0 & \sigma^2_d \end{bmatrix} \tag{2.2.5}$$

2.3 State space representation

This section provides the derivation of transition and measurement equations, which I will need to filter the hidden variables: expected returns and expected dividend growth. Again, this section follows closely Binsbergen and Koijen (2010).

Binsbergen and Koijen (2010)define demeaned expected returns and dividend growth as

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

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

Then equations (2.2.2) and (2.2.3) - transition equations - are equivalent to:

$$\hat{\mu}_{t+1} = \delta_1 \hat{\mu}_t + \epsilon^\mu_{t+1} \tag{2.3.1}$$

$$\hat{g}_{t+1} = \gamma_1 \hat{g}_t + \epsilon^g_{t+1} \tag{2.3.2}$$

The measurement equations are given by:

$$\Delta d_{t+1} = \gamma_0 + \hat{g}_t + \epsilon^d_{t+1} \tag{2.3.3}$$

$$p_t - d_t = A - B_1 \hat{\mu}_t + B_2 \hat{g}_t \tag{2.3.4}$$

Let me define $pd_t = p_t - d_t$.

Now, I iterate 1 step forward the equation (2.3.4) to get

$$\begin{aligned} pd_{t+1} &= A - B_1 \hat{\mu}_{t+1} + B_2 \hat{g}_{t+1} \\ &= A - B_1(\delta_1 \hat{\mu}_t + \epsilon^{\mu}_{t+1}) + B_2(\gamma_1 \hat{g}_t + \epsilon^g_{t+1}) \\ &= A - B_1 \delta_1 \hat{\mu}_t - B_1 \epsilon^{\mu}_{t+1} + B_2 \gamma_1 \hat{g}_t + B_2 \epsilon^g_{t+1} \\ &= A + \delta_1(-B_1 \hat{\mu}_t) + B_2 \gamma_1 \hat{g}_t - B_1 \epsilon^{\mu}_{t+1} + B_2 \epsilon^g_{t+1} \\ &= A + \delta_1(pd_t - A - B_2 \hat{g}_t) + B_2 \gamma_1 \hat{g}_t - B_1 \epsilon^{\mu}_{t+1} + B_2 \epsilon^g_{t+1} \\ &= (1 - \delta_1)A + B_2(\gamma_1 - \delta_1)\hat{g}_t + \delta_1 pd_t - B_1 \epsilon^{\mu}_{t+1} + B_2 \epsilon^g_{t+1} \end{aligned} \tag{2.3.5}$$

The **transition equation** is

$$\hat{g}_t = \gamma_1 \hat{g}_{t-1} + \epsilon^g_t \tag{2.3.6}$$

The **measurement equation 1** is

$$\Delta d_{t+1} = \gamma_0 + \hat{g}_t + \epsilon^d_{t+1} \tag{2.3.7}$$

The **measurement equation 2** is

$$pd_{t+1} = (1 - \delta_1)A + B_2(\gamma_1 - \delta_1)\hat{g}_t + \delta_1 pd_t - B_1 \epsilon^{\mu}_{t+1} + B_2 \epsilon^g_{t+1} \tag{2.3.8}$$

In the next chapter, I will transform the equations 2.3.6,2.3.7,2.3.8 into standard space form using matrix algebra as in Binsbergen and Koijen (2010).

Chapter 3

Kalman Filter and Sequential Monte Carlo

In this chapter, I derive the Kalman filter for present value model with latent variable and explain sequential Monte Carlo algorithm (SMC). Both Kalman filter and SMC will be used to estimate posterior parameters of the model.

3.1 Kalman Filter: Bayesian Derivation and Application to Present Value Model

In this section, I derive Kalman filter in a Bayesian way. My explanation and notation are based on Simo Sarkka's lecture "Bayesian Optimal Filtering Equations and Kalman Filter" and on derivation in the article by Binsbergen and Koijen (2010).

3.1.1 Standard State-Space Form

I write the equations (2.3.6), (2.3.7), (2.3.8) in the standard state space form to account for the time lag in the initial state space form:

$$\begin{aligned} X_t &= FX_{t-1} + \Gamma \epsilon_t^X \\ Y_t &= M_0 + M_1 Y_{t-1} + M_2 X_t \end{aligned}$$

where

$$X_t = \begin{bmatrix} \hat{g}_{t-1} \\ \epsilon_t^d \\ \epsilon_t^g \\ \epsilon_t^\mu \end{bmatrix}, \quad Y_t = \begin{bmatrix} \Delta d_t \\ pd_t \end{bmatrix}$$

$$F = \begin{bmatrix} \gamma_1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \epsilon_t^X = \begin{bmatrix} \epsilon_t^d \\ \epsilon_t^g \\ \epsilon_t^\mu \end{bmatrix}$$

with variance of ϵ_t^X defined in (2.2.5). The measurement equation matrices are

$$\begin{aligned} M_0 &= \begin{bmatrix} \gamma_0 \\ (1-\delta_1)A \end{bmatrix} \\ M_1 &= \begin{bmatrix} 0 & 0 \\ 0 & \delta_1 \end{bmatrix} \\ M_2 &= \begin{bmatrix} 1 & 1 & 0 & 0 \\ B_2(\gamma_1 - \delta_1) & 0 & B_2 & -B_1 \end{bmatrix} \end{aligned}$$

.

In probabilistic terms

$$
\begin{aligned}
p(X_t|X_{t-1}) &= \mathcal{N}(X_t \mid F_{t-1}X_{t-1},\, \Gamma\Sigma\Gamma^T) \\
Y_t &= M_0 + M_1 Y_{t-1} + M_2 X_t
\end{aligned}
$$

3.1.2 Kalman Filter

Prediction Step

In this step, I want to come up with the prediction of X_t given $Y_{1:t-1}$. Given information now, I want to infer the expectations for the next period about the dividend growth rate and returns. First, assume that the posterior of the previous step is Gaussian

$$
p(X_{t-1} \mid Y_{1:t-1}) = \mathcal{N}\left(X_{t-1} \mid m_{t-1},\, P_{t-1}\right)
$$

Note the following

$$
\begin{aligned}
p(X_t,\, X_{t-1}|Y_{1:t-1}) &= p(X_t \mid X_{t-1},\, Y_{1:t-1}) \times p(X_{t-1} \mid Y_{1:k-1}) \\
&= p(X_t \mid X_{t-1},) \times p(X_{t-1} \mid Y_{1:t-1}) \\
&= \underbrace{\mathcal{N}(x_k|F_{t-1}X_{t-1},\, \Gamma\Sigma\Gamma_k^T)}_{\text{Transition equation}} \times \underbrace{\mathcal{N}(X_{t-1}|m_{t-1},\, P_{t-1})}_{\text{Posterior from prev step}} \\
&= \mathcal{N}\left(\begin{bmatrix} X_{t-1} \\ X_t \end{bmatrix} \middle| \begin{bmatrix} m_{t-1} \\ Fm_{t-1} \end{bmatrix},\, \begin{bmatrix} P_{t-1} & FP_{t-1} \\ FP_{t-1} & FP_{t-1}F^T + \Gamma\Sigma\Gamma^T \end{bmatrix} \right)
\end{aligned}
$$

Next, by integrating over X_{t-1}, I get the following equation

$$
\begin{aligned}
p(X_t|Y_{1:t-1}) &= \int p(X_t,\, X_{t-1}|Y_{1:t-1})dX_{k-1} \\
&= \int p(X_t \mid X_{t-1},\, Y_{1:t-1}) \times p(X_{t-1} \mid Y_{1:t-1})dX_{t-1} \\
&= \int p(X_t \mid X_{t-1}) \times p(X_{t-1} \mid Y_{1:t-1})dX_{t-1} \\
&= \int \mathcal{N}(X_t|FX_{t-1},\, \Gamma\Sigma\Gamma^T) \times \mathcal{N}(X_{t-1}|m_{t-1},\, P_{t-1})dX_{t-1} \\
&= \underbrace{\mathcal{N}(X_t|\, Fm_{t-1},\, FP_{t-1}F^T + \Gamma\Sigma\Gamma^T)}_{\text{Marginalization}}
\end{aligned}
$$

Denote

$$
\begin{aligned}
m_t^- &= Fm_{t-1} \\
P_t^- &= FP_{t-1}F^T + \Gamma\Sigma\Gamma^T
\end{aligned}
$$

Update Step

The joint distribution $p(X_t,\, Y_t|Y_{1:t-1})$ is

$$
p(X_t, Y_t|Y_{1:t-1}) = \mathcal{N}\left(\begin{bmatrix} X_t \\ Y_t \end{bmatrix} \mid \begin{bmatrix} m_t^- \\ M_0 + M_1Y_{t-1} + M_2m_t^- \end{bmatrix},\, \begin{bmatrix} P_t^- & P_t^-M_2^T \\ M_2P_t^- & M_2P_t^-M_2^T \end{bmatrix} \right)
$$

Then the marginal distribution $p(Y_t|Y_{1:t-1})$ can be computed

$$
p(Y_t|Y_{1:t-1}) = \mathcal{N}\left(Y_t|M_0 + M_1Y_{t-1} + M_2m_t^-,\, M_2P_t^-M_2^T \right)
$$

The conditional distribution $p(X_t|Y_{1:t})$ is then given by

$$
p(X_t|Y_{1:t}) = \mathcal{N}\left(X_t \mid m_t,\, P_t \right)
$$

where

$$\begin{aligned} m_t &= m_t^- + P_t^- M_2^T \underbrace{(M_2 P_t^- M_2^T)^{-1}}_{S_t^{-1}} \underbrace{(Y_t - M_0 - M_1 Y_{t-1} - M_2 m_t^-)}_{v_t} \\ P_t &= P_t^- - P_t^- M_2^T \underbrace{(M_2 P_t^- M_2^T)^{-1}}_{S_t^{-1}} M_2 P_t^- \end{aligned}$$

Denote

$$\begin{aligned} S_t &= M_2 P_t^- M_2^T \\ K_t &= P_t^- M_2^T S_t^{-1} \\ v_t &= Y_t - M_0 - M_1 Y_{t-1} - M_2 m_t^- \end{aligned}$$

Note

$$S_t K_t^T = S_t S_t^{-1} M_2 P_t^- = M_2 P_t^-$$

Then I can write m_t and P_t as

$$\begin{aligned} m_t &= m_t^- + K_t v_t \\ P_t &= P_t^- - K_t S_t K_t^T \end{aligned}$$

To sum up, the equations of the Kalman filter are

- ⋆ **Initialization**:

$$\begin{aligned} m_0 &= 0_{1\times 4} \\ P_0 &= \begin{bmatrix} \mathrm{Var}(\hat{g}) & 0 & 0 & 0 \\ 0 & \sigma_\mu^2 & 0 & 0 \\ 0 & 0 & \sigma_g^2 & 0 \\ 0 & 0 & 0 & \sigma_d^2 \end{bmatrix} \end{aligned}$$

where $\mathrm{Var}(\hat{g})$ is the variance of stationary distribution of $\hat{g}_t$

⋆ **Prediction step**:

$$\begin{aligned} m_t^- &= Fm_{t-1} \\ P_t^- &= FP_{t-1}F^T + \Gamma\Sigma\Gamma^T \end{aligned}$$

⋆ **Update step**:

$$\begin{aligned} S_t &= M_2 P_t^- M_2^T \\ K_t &= P_t^- M_2^T S_t^{-1} \\ v_t &= Y_t - M_0 - M_1 Y_{t-1} - M_2 m_t^- \\ m_t &= m_t^- + K_t v_t \\ P_t &= P_t^- - K_t S_t K_t^T \end{aligned}$$

Maximum Likelihood

Suppose number of periods is T. Then the conditional log-likelihood of observing Y_t where $t = 1, ..., T$ given $Y_{1:t-1}$ have occurred can be computed as

$$\log \mathcal{L}_t \propto -v_t^T S_t^{-1} v_t - \log(\det(S_t))$$

Then the likelihood of observing the whole sample $Y_{1:T}$ is

$$\log \mathcal{L}_{1:T} = \sum_{k=1}^{T} \log \mathcal{L}_t$$

Computing Filtered Series

The Kalman filter produces m_t which has $E[\hat{g}_{t-1}|Y_{1:T}]$ as the first component and $E[\epsilon_t^d|Y_{1:T}]$ $E[\epsilon_t^g|Y_{1:T}]$ $E[\epsilon_t^\mu|Y_{1:T}]$ as the other three components. However, I need $E[\hat{g}_t|Y_{1:T}]$ and $E[\hat{\mu}_{t-1}|Y_{1:T}]$

which can be computed using m_t, (2.3.1),(2.3.2)and(2.3.4).

I compute the series for the filtered demeaned expected return as

$$\begin{aligned} E_t[\hat{\mu}_t] &= E_t[\delta_1 \hat{\mu}_{t-1} + \epsilon_t^{\mu}] \\ &= E_t\left[\delta_1 \left(\frac{A + B_2 \hat{g}_{t-1} - pd_t}{B_1}\right) + \epsilon_t^{\mu}\right] \\ &= \delta_1 \frac{A}{B_1} + \delta_1 \frac{B_2}{B_1} E_t[\hat{g}_{t-1}] - \frac{\delta_1}{B_1} pd_t + E_t[\epsilon_t^{\mu}] \end{aligned}$$

and the series for the filtered demeaned expected dividend growth rate as

$$E_t[\hat{g}_t] = \gamma_1 E_t[\hat{g}_{t-1}] + E_t[\epsilon_t^g]$$

where $E_t[\hat{\mu}_t]$=$E[\hat{\mu}_t|Y_{1:T}]$ and $E_t[\hat{g}_t] = E[\hat{g}_t|Y_{1:T}]$. Then I compute filtered series $E_t[\mu_t] = E_t[\hat{\mu}_t]$+$\delta_0$ and $E_t[g_t] = E_t[\hat{g}_t] + \gamma_0$ which I use to forecast Δd_{t+1} and r_{t+1}, respectively.

3.2 Sequential Monte Carlo

This note explains the Sequential Monte Carlo (SMC) algorithm which is a more robust way to estimate parameters than maximum likelihood. SMC accounts for parameter uncertainty: it treats a parameter vector as random vector and computes its distribution for each time period. To compare, in maximum likelihood, parameters are perceived as fixed and unknown numbers. The explanations in this section are based on Chopin (2002).

The posterior density is given by:

$$\gamma_t(\theta \mid y_{1:t}) \propto f(y_{1:t} \mid \theta) \times f(\theta) \, \text{for } t = 1, ..., T$$

where $f(y_{1:t} \mid \theta)$ is a likelihood and $f(\theta)$ is a prior. The log of posterior (target distribution) is then

$$\log\Big(\gamma_t(\theta \mid y_{1:t})\Big) \propto \log\Big(f(y_{1:t} \mid \theta)\Big) + \log\Big(f(\theta)\Big)$$

Assume $\{y\}_t$ is a Markov process of order 1, that is current observation depends only on the previous observation. Then the incremental weight of a particle $\theta^{(n)}$ at time $t+1$ is defined as

$$\begin{aligned}
\frac{\gamma_{t+1}(\theta \mid y_{1:t+1})}{\gamma_t(\theta \mid y_{1:t})} &\propto \frac{f(y_{1:t+1} \mid \theta) \times f(\theta)}{f(y_{1:t} \mid \theta) \times f(\theta)} \\
&= \frac{f(y_{1:t+1} \mid \theta)}{f(y_{1:t} \mid \theta)} \\
&= f(y_{t+1} \mid y_{1:t},\, \theta) \\
&= f(y_{t+1} \mid y_t,\, \theta)
\end{aligned}$$

In sequential Monte Carlo, the weight of each parameter $\theta^{(n)}$ vector (where $n = 1, ..., N$) at time t is given by the recursion

$$w_t^{(n)} = f(y_t \mid y_{t-1},\, \theta_{t-1}^{(n)}) \times w_{t-1}^{(n)}$$

The sequential Monte Carlo algorithm is given by the following steps

Initialization: Generate a cloud of N weighted particles $\{\theta_0^{(n)},\, w_0^{(n)}\}_{n=1}^{N}$ representing posterior $\gamma_0(\theta)$.

- ⋆ Assume prior $f(\theta)$
- ⋆ Given prior $f(\theta)$, generate N parameter vectors $\theta_0^{(n)}$ where $n = 1, ..., N$
- ⋆ Compute prior log-likelihood of observing each parameter vector $\theta_0^{(n)}$ for $n = 1, .., N$ given the prior $f(\theta)$
- ⋆ For each generated parameter vector, set the hidden state vector equal zero
- ⋆ For each generated parameter vector, initialize hidden state variance-covariance matrix the same way as for Kalman filter.

⋆ Set log-likelihood and log-weights equal to zero for each parameter vector.

Loop: For $t = 1, .., T$ get the distribution $\{\theta_t^{(n)},\, w_t^{(n)}\}_{n=1}^N$ representing the posterior $\gamma_t(\theta)$ in the following way

⋆ For each $\theta^{(n)}$ (where $n = 1, .., N$), compute the incremental log-weight $\log\Big(f(y_t|y_{t-1},\, \theta_t^{(n)})$ by using the Kalman filter

⋆ For each $\theta^{(n)}$ update the log-weights

$$\log(w_t^{(n)}) = \underbrace{\log\Big(f(y_t \mid y_{t-1},\, \theta_{t-1}^{(n)})\Big)}_{\text{log of incremental weight}} + \log(w_{t-1}^{(n)})$$

⋆ For each $\theta^{(n)}$, compute the normalized weights $\pi_t^{(n)}$

$$\pi_t^{(n)} = \frac{w_t^{(n)}}{\sum_{n=1}^N w_t^{(n)}}$$

⋆ For each $\theta^{(n)}$, compute Effective Sample Size (ESS) defined as

$$\text{ESS} = \frac{1}{\sum_{n=1}^N (\pi_t^{(n)})^2}$$

⋆ IF ESS$> B$

♣ keep $\{\theta_t,\, \pi_t\}_{n=1}^N$

⋆ ELSE

♣ re-sample $\{\theta_t^{(n)}\}_{n=1}^N$

♣ set $\pi_t^{(n)} = \dfrac{1}{N}$

♣ WHILE number of unique particles is less than C or no move step has been done, do the following move step for each particle $n = 1, ..., N$

- Sample $u_t^{(n)} \sim U_{[0,1]}$ where U is a uniform distribution
- Sample $\theta_t^{(n)*} \sim h_t(\cdot)$ where $h(\cdot)$ is an independent normal
- IF $\log u_t^{(n)} < \log \alpha = \min \left\{ 0, \log \gamma_t(\theta_t^{(n)*}) - \log \gamma_t(\theta_t^{(n)}) + \log h_t(\theta_t^{(n)}) - \log h_t(\theta_t^{(n)*}) \right\}$

  · then $\theta_t^{(n)} = \theta_t^{(n)*}$

- ELSE

  · $\theta_t^{(n)} = \theta_t^{(n)}$.

In the next chapters I present data and its estimation using Kalman filter and sequential Monte Carlo method

Chapter 4

Data

4.1 Data

I look at at the period 1946-2013. I get the nominal time series for S&P composite prices and dividend series from Shiller's website. I get monthly interest rate for 3 month treasury bill from Federal Reserve at St Louis. Then, I adjust the dividend time series for the monthly interest rate of 3-month T-bill to get cash-invested dividends. After that, I construct annual data for stock prices by taking the price at the end of the year and annual data for cash-invested dividends by taking the average for the whole year. Further, I modify the annual series for stock-prices and for cash-invested dividends to obtain annual series for price-dividend ratio and dividend growth rate. Below, the table 4.1 summarizes the statistics for the dividend growth rate, price-dividend ratio and the interest rate.

Table 4.1 – Summary Statistics

	Δd	pd	r
Mean	0.0580	3.4651	0.1049
Median	0.0521	3.4101	0.1397
Standard Deviation	0.0639	0.4419	0.1598
Maximum	0.2087	4.4428	0.3840
Minimum	-0.1286	2.7341	-0.4900

Next, I plot dividend growth rate, price dividend ratio and the interest rate against the time.

Figure 4.1 – Dividend Growth Rate

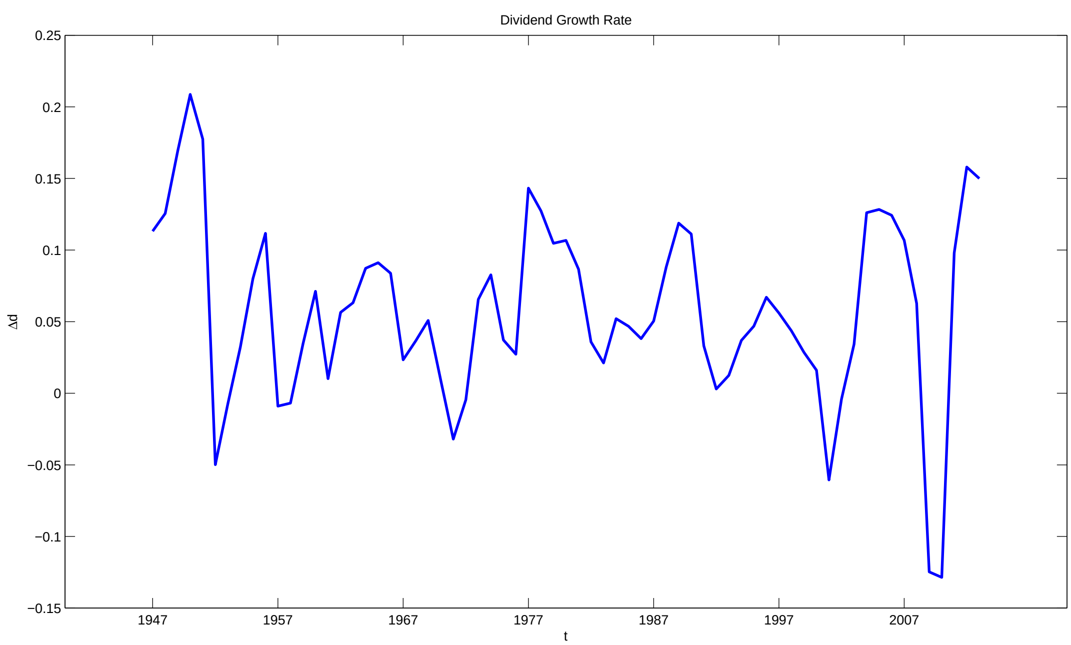


Figure 4.2 – Price-Dividend Ratio: Reinvesting in Risk Free Rate

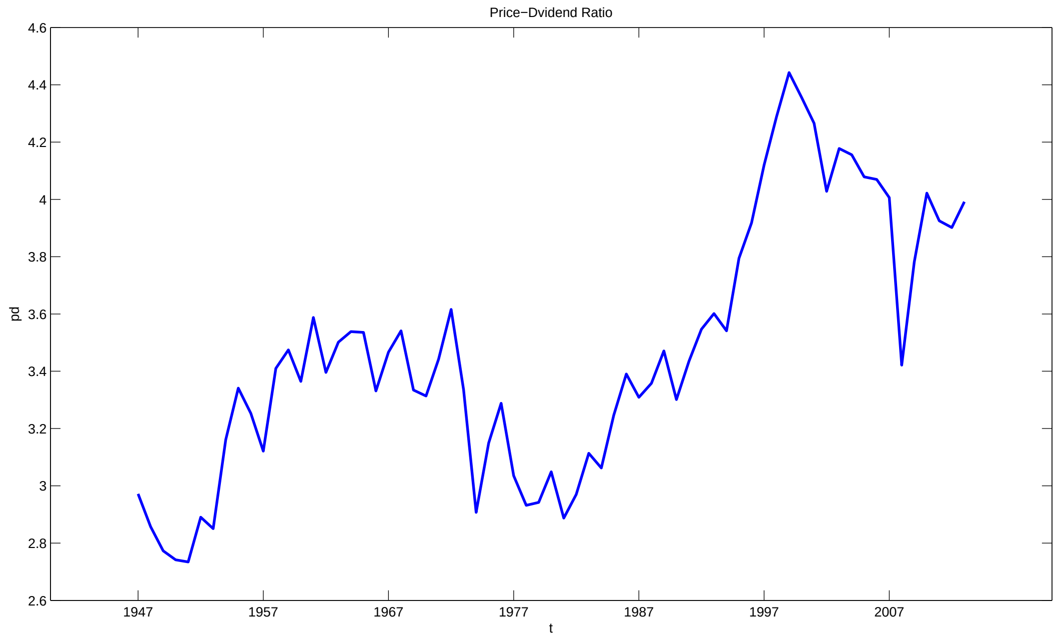
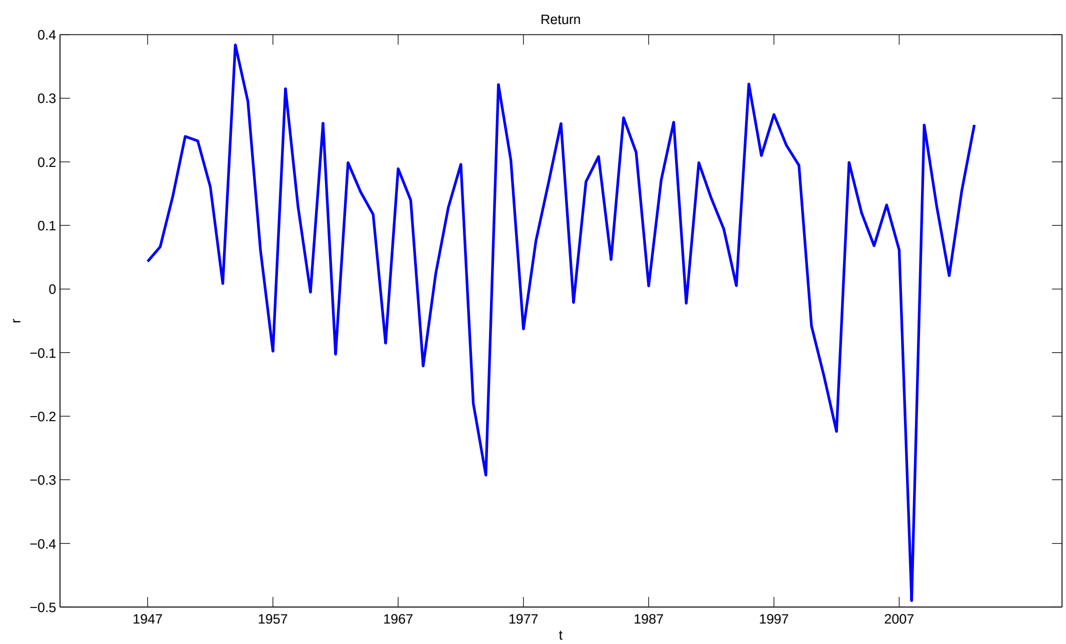


Figure 4.3 – Return



These graphs are similar to the ones in Binsbergen and Koijen (2010). .

As the next step to check that my data is correct, I run OLS predictive regressions

$$\begin{aligned} \triangle d_{t+1} &= \alpha_d + \beta_d pd_t + \varepsilon_{t+1} \\ r_{t+1} &= \alpha_r + \beta_d pd_t + \varepsilon_{t+1} \end{aligned}$$

The table 4.2 and table 4.3 show the results.

Table 4.2 – OLS Prediction of r_{t+1} with pd_t

	Coefficient	t-statistics	P-value
β_r	-0.12	-2.74	0.008

$R^2 = 0.1$

Table 4.3 – OLS Prediction of $\Delta d_{t+1}$with pd_t

	Coefficient	t-statistics	P-value
$\beta_{\Delta d}$	-0.014	0.75	0.454

$R^2 = 0.01$

The tables above show that β_r is negative and statistically significant while $\beta_{\Delta d}$ is not statistically significant. This corresponds to observations pointed out by Cochrane (2005).Given plots of series and the results of the predictive OLS, I conclude that the data is suitable for further investigation.

Chapter 5

Simulations

In this chapter, I present the results from Kalman filter and sequential Monte Carlo algorithm on simulated data to make sure that algorithms work properly.

5.1 Kalman Filter Simulation Results

To check that Kalman filter works well, I first generate data $\hat{g}_t$, Δd_t and pd_t using the transition equation 2.3.6 and measurement equations 2.3.7 and 2.3.8. The parameters of data generating process are presented in the following table

Table 5.1 – Parameters Used in DGP for Simulation

γ_0	γ_1	δ_0	δ_1	σ_d	σ_g	σ_μ
0.062	0.354	0.090	0.932	0.02	0.03	0.016

First, I plot the simulated demeaned expected dividend growth $\hat{g}_t$and filtered expected demeaned dividend growth $E_t[\hat{g}_t]$.

Figure 5.1 – True De-meaned Expected Dividend Growth$\hat{g}_t$ and Filtered One $E_t[\hat{g}_t]$

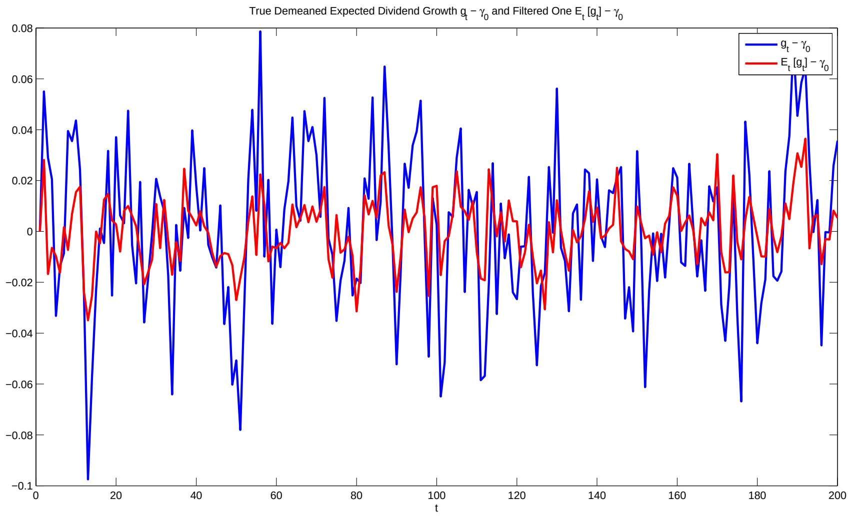
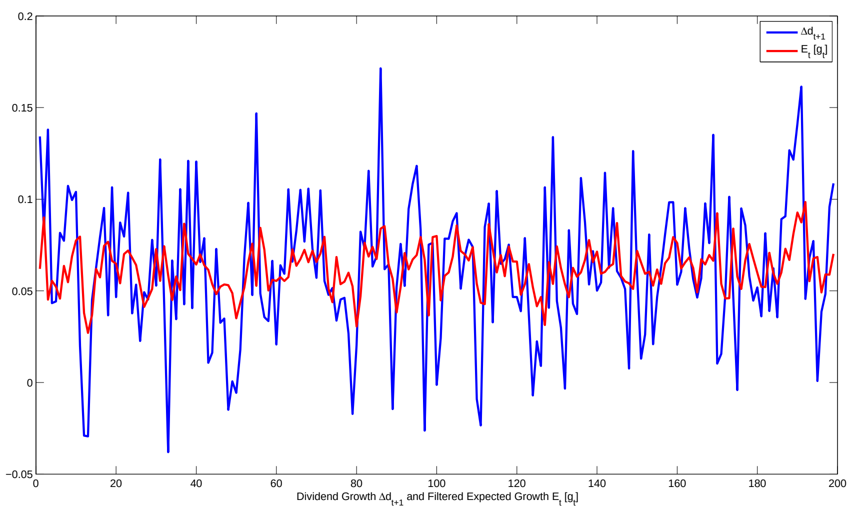


Figure 5.1 shows that the filtered series for the demeaned dividend growth is close to the true one. Second, I plot the series for dividend growth $\{\triangle d_{t+1}\}_{t=1}^{T-1}$ and the filtered series of expected dividend growth $\{E_t[g_t]\}_{t=1}^{T-1}$.

Figure 5.2 – Dividend Growth Δd_{t+1} and Filtered Expected Dividend Growth $E_t[g_t]$



Figures 5.2 shows that the filtered series for expected dividend growth tracks the dividend growth series quite well. Next, I present the log likelihood plots to make sure that true parameters have the highest likelihood.

Figure 5.3 – Log-likelihood Profile Plots for γ_0,γ_1, δ_0, and δ_1

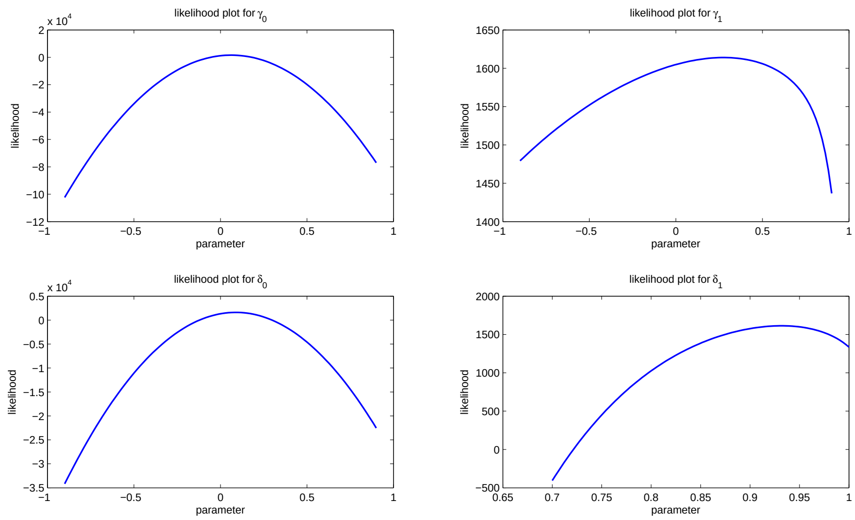
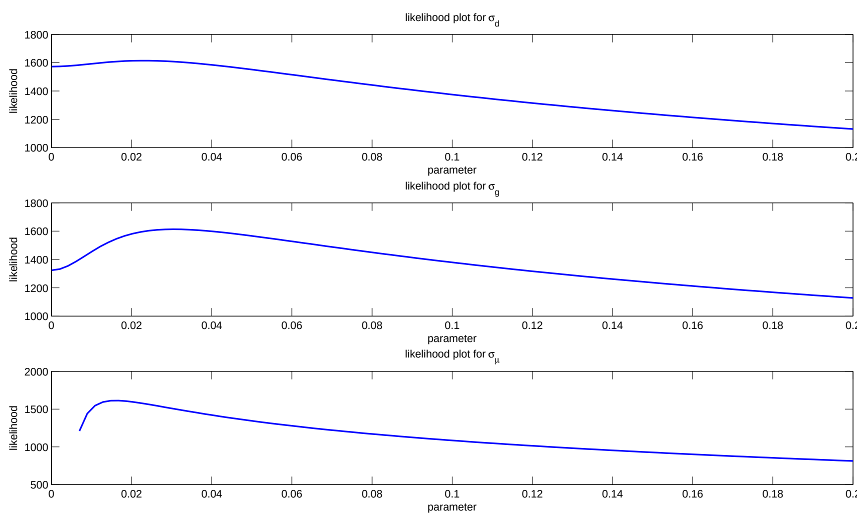


Figure 5.4 – Log-likelihood Profile Plots for σ_d, σ_g, σ_μ



The likelihood profile plots have peaks at the true parameter value. I conclude therefore that the Kalman filter works properly.

5.2 Sequential Monte Carlo Simulation Results

In this section I present the results of sequential Monte Carlo simulation. In the simulation, the number of periods is 300, the number of parameter vectors per 1 time period is 2000 and I trigger move-step if ESS is less than $0.4 * number\,of\,parameters$. Table 5.2 shows the true and posterior parameter estimates.

Table 5.2 – Posterior Parameter Estimates

	γ_0	γ_1	δ_0	δ_1	σ_d	σ_g	σ_μ
True	0.062	0.354	0.090	0.932	0.02	0.03	0.016
Posterior	0.0575	0.3440	0.0966	0.9347	0.0200	0.0307	0.0156

The posterior parameter estimates are close to true parameter estimates.

Next in Figures 5.5 and 5.6, I show the plots of parameters to check for convergence to true values in as the number of periods increases.

Figure 5.1 – Posterior Estimates of γ_0, γ_1, δ_0, δ_1

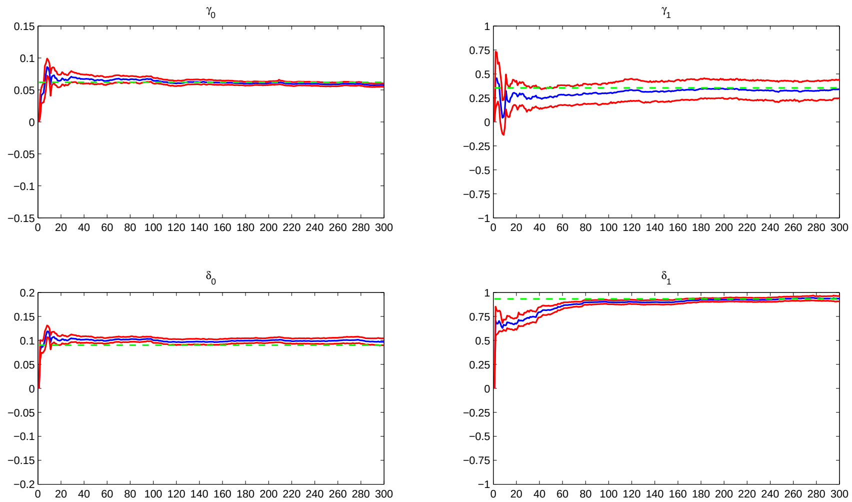
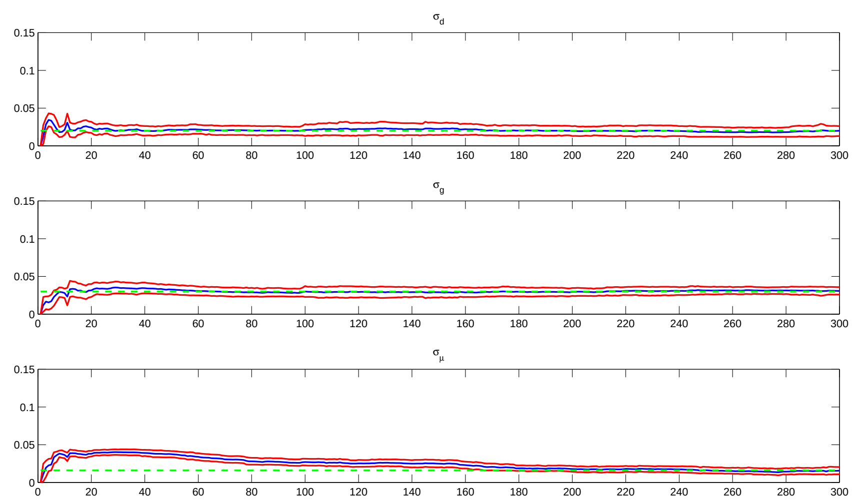


Figure 5.2 – Posterior Estimates of σ_d, σ_g, σ_μ



Figures above show that variance for all parameters decreases pushing parameter estimates to true values except for γ_1. In case of γ_1, the mean of distribution at $T = 300$ coincides with true value; however, I cannot see that the variance decreases as time progresses. This observation corresponds to high standard errors of γ_1 when Binsbergen and Koijen (2010)estimate γ_1 with maximum likelihood.

Chapter 6

Results

In this chapter, I present the results of sequential Monte Carlo estimation. In addition, I do model comparison.

6.1 SMC Estimation Results

I run sequential Monte Carlo procedure with for the data described in chapter 4. The number of parameter vectors for each time period is 2500. The next two figures present the results of 5 runs of SMC. Solid line stands for the mean, dashed lines stand for 5th and 95th percentiles.

Figure 6.1 – SMC Results for γ_0, γ_1, δ_0, δ_1

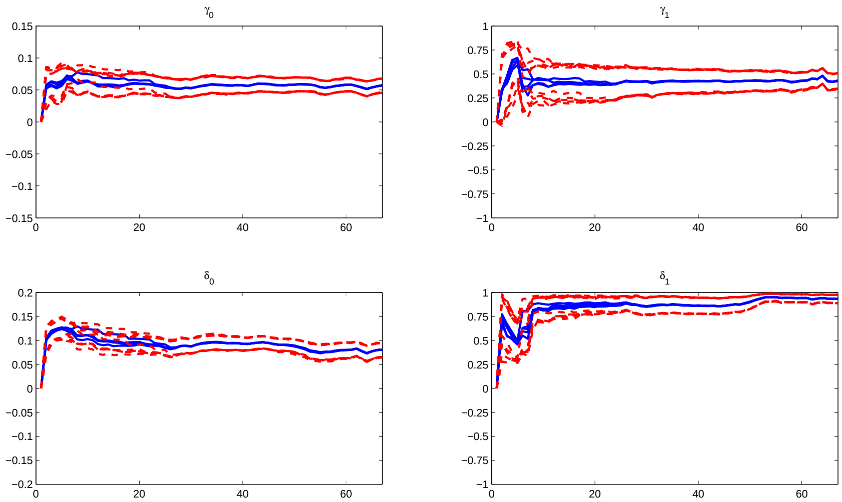
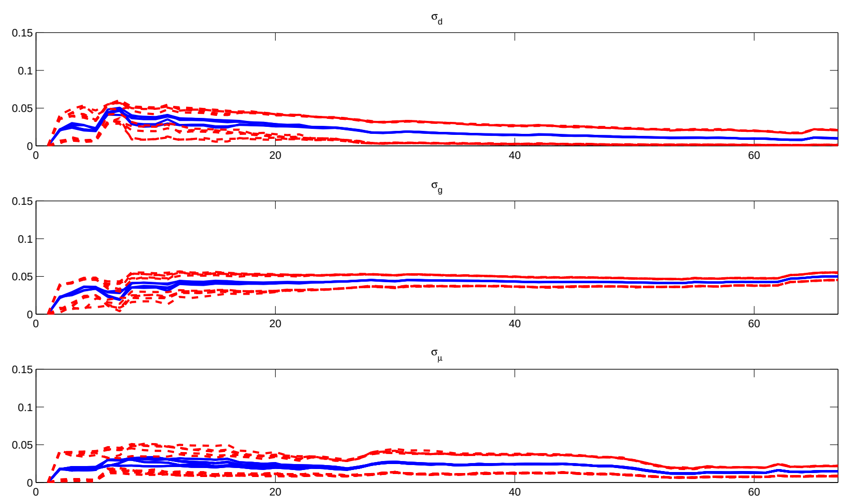


Figure 6.2 – SMC Results for σ_d, σ_g, σ_μ



The figures 6.1 and 6.2 show that parameters converge. Table 6.1 provides the mean of posterior estimates of 5 runs and the standard deviation. The figures above and the table below show the posterior estimates are close to each other in all 5 runs.

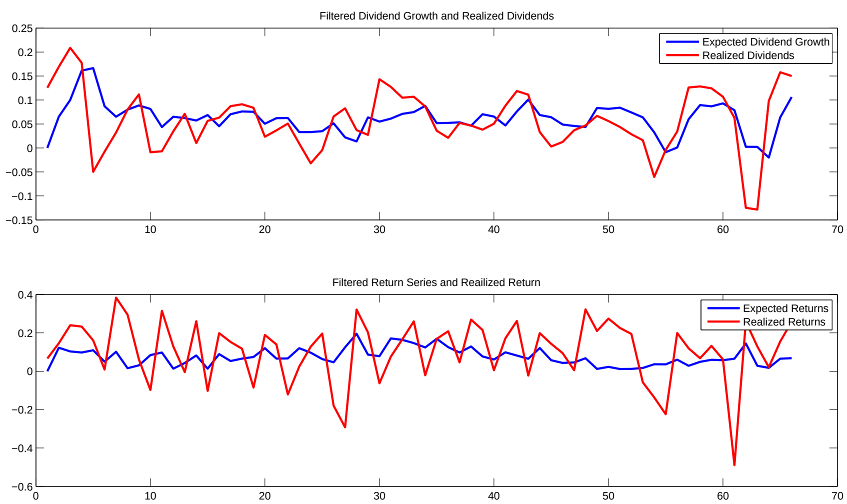
Table 6.1 – SMC Parameter Estimates, Nominal Data

	γ_0	γ_1	δ_0	δ_1	σ_d	σ_g	σ_μ
Mean	0.0572	0.4287	0.0805	0.9335	0.0098	0.0500	0.0148
St.d.	0.0002	0.0019	0.0003	0.0005	0.0002	0.0001	0.0001

Table 6.1 also shows that the estimates are close to the ones in Binsbergen and Koijen (2010).

Next, I am going to plot the the series of realized dividend growth against the filtered expected dividend growth (that is, $\triangle d_{t+1}$ against $E_t[g_t]$) and the series of the realized return against filtered expected return (that is, r_{t+1} against $E_t[\mu_t]$) of taking the average of 5 runs. Figure 6.3 shows the series produced by sequential Monte Carlo.

Figure 6.3 – SMC Out-of-Sample: Series Δd_{t+1}against filtered g_t and r_{t+1}against filtered μ_t



If I run predictive OLS regressions

$$\begin{aligned} \triangle d_{t+1} &= \alpha_d + \beta_d E_t[g_t] + \epsilon_{t+1} \\ r_{t+1} &= \alpha_r + \beta_r E_t[\mu_t] + \epsilon_{t+1} \end{aligned}$$

then I get the following results summarized in the tables below

Table 6.2 – Real-Time: Predicting r_{t+1} with $E_t[\mu_t]$

$R^2 = 0.05$	Coefficient	t-stat	P-value
β_r	0.82	1.88	0.06

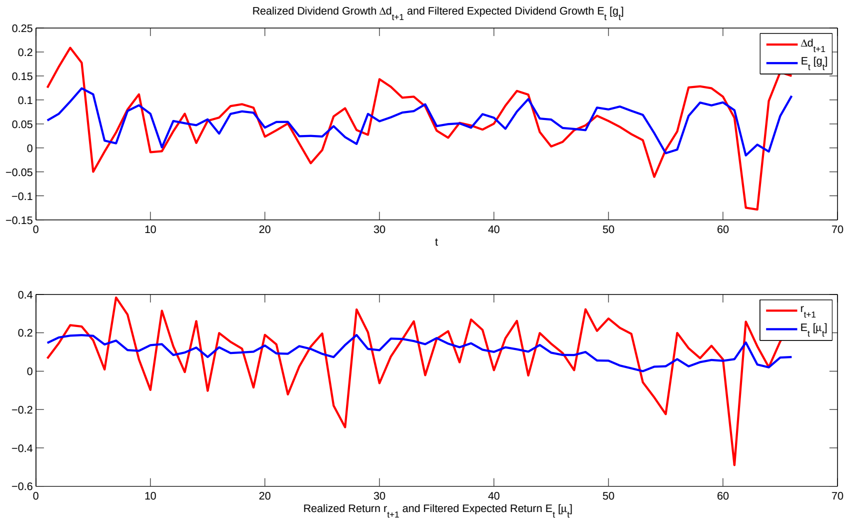
Table 6.3 – Real-Time: Predicting $\triangle d_{t+1}$ with $E_t[g_t]$

$R^2 = 0.14$	Coefficient	t-stat	P-value
$\beta_{\Delta d}$	0.74	3.26	0.002

The conclusion from the Table 6.3: If I use Bayesian learning (SMC) in real time, then dividend growth is better predictable than return as the tables above show. The slope coefficient of the OLS for returns has a P-value higher than 0.05. This confirms that returns are hard to predict and thus explains why stocks have high risk premium.

In SMC, I use only the present information to predict future dividend growth and return. SMC and OLS show how well filtered series of expected dividend growth and return perform out-of-sample. This corresponds to predicting in real time. To compare, in maximum likelihood I use the information from the future to predict the dividend growth and returns. This corresponds to predicting in-sample. To see how well the filtered series predict in-sample , I run the Kalman filter using the data and the parameter vector from the Table 6.1. The Figure 6.4 shows the data and the filtered series.

Figure 6.4 – KF In-Sample: Series Δd_{t+1}against filtered g_t and r_{t+1}against filtered μ_t



The Tables 6.4 and 6.5 show the results when I run r_{t+1} and $\triangle d_{t+1}$ on $E_t[\mu_t]$ and $E_t[g_t]$ in-sample, that is, using the Kalman filter, already optimal parameter vector and the the information from the future.

Table 6.4 – In-Sample: Predicting r_{t+1} with $E_t[\mu_t]$

$R^2 = 0.14$	Coefficient	t-stat	P-value
β_r	1.08	2.69	0.009

Table 6.5 – In-Sample: Predicting $\triangle d_{t+1}$ with $E_t[g_t]$

$R^2 = 0.33$	Coefficient	t-stat	P-value
$\beta_{\Delta d}$	1.17	5.64	0.000

As expected, returns and dividend growth are better predicted in-sample than out-of-sample. The Tables 6.4 and 6.5 above with R^2 and slope coefficients and t-statistic confirm that.

The comparison of the Bayesian learning (out-of-sample) forecasting and the Kalman filter (in-sample forecasting) reveals that there is a bias when forecasting returns and the dividend growth in-sample. Moreover, in-sample forecast are more optimistic than out-of-sample forecasts because the slope coefficients for the Kalman Filter filtered series are higher than the slope coefficients for the SMC filtered series.

6.2 Model Comparison

In addition to the out-of-sample and in-sample forecasting, I can do the model comparison using the sequential Monte Carlo method. First, to do the model comparison, let me define the incremental normalizing ratio as

$$\begin{aligned}
\frac{Z_t}{Z_{t-1}} &= \frac{\int_\Theta \gamma_t(\theta) d\theta}{\int_\Theta \gamma_{t-1}(\theta) d\theta} \\
&= \frac{\int_\Theta f(y_{1:t} \mid \theta) \times f(\theta) d\theta}{\int_\Theta f(y_{1:t-1} \mid \theta) \times f(\theta) d\theta} \\
&= \frac{f(y_{1:t})}{f(y_{1:t-1})} \\
&= f(y_t | y_{1:t-1})
\end{aligned}$$

Also, the incremental normalizing ratio can be written as

$$\frac{Z_t}{Z_{t-1}} = \int_\Theta \frac{\gamma_t(\theta)}{\gamma_{t-1}(\theta)} \frac{\gamma_{t-1}(\theta)}{Z_{t-1}} d\theta$$

In practice, given a cloud of particles $\{\theta_{t-1}^{(n)}, w_{t-1}^{(n)}\}_{n=1}^N$, the incremental normalizing ratio is computed as

$$\frac{\hat{Z_t}}{Z_{t-1}} = \sum_{n=1}^{N}\left(\frac{w_t^{(n)}}{w_{t-1}^{(n)}} \times \frac{w_{t-1}^{(n)}}{\sum_{n=1}^{N} w_{t-1}^{(n)}}\right)$$

$$= \sum_{n=1}^{N}\left(\underbrace{f(y_t \mid y_{t-1},\, \theta_{t-1}^{(n)})}_{\text{incremental weight}} \times \overbrace{\frac{w_{t-1}^{(n)}}{\sum_{n=1}^{N} w_{t-1}^{(n)}}}^{\text{normalized weight}}\right)$$

$$= \sum_{n=1}^{N}\left(\underbrace{f(y_t \mid y_{t-1},\, \theta_{t-1}^{(n)})}_{\text{incremental weight}} \times \pi_{t-1}^{(n)}\right)$$

Second, let me define 3 models based on the equations 2.3.6,2.3.7 and2.3.8.

- ⋆ Model 1: the model with all parameters, that is, with γ_0,γ_1, δ_0, δ_1, σ_d, σ_g, σ_μ
- ⋆ Model 2: the model without γ_1
- ⋆ Model 3: the model without δ_1

Given the data and SMC results, it is possible to compare the three models, which I leave for the future research.

Chapter 7

Conclusion

Sequential Monte Carlo and the present value model can be used for both in-sample forecasting and out-of-sample forecasting. The in-sample forecasts are more optimistic than the out-of-sample forecasts. Out-of-sample, dividend growth rate is predictable with $R^2 = 0.14$ and returns are close to being unpredictable with slope coefficient on the verge of being statistically insignificant. This implies that returns may follow the martingale process. In-sample, things are better: both return and dividend growth are predictable with $R^2_{ret} = 0.14$ and $R^2_{div} = 0.33$. This shows that in-sample forecasting is overoptimistic and should be taken with caution.

Bibliography

Binsbergen, J. and R. Koijen. 2010. "Predictive Regressions: A Present Value Approach." *Journal of Finance* 65:1439–1471.

Chopin, N. 2002. "A Sequential Particle Filter Method for Static Models." *Biometrica* 89:539–551.

Cochrane, J. 2008. "The Dog That Did Not Bark: A Defense of Return Predictability." *The Review of Financial Studies* 21:1534–1575.

Cochrane, J.H. 2005. *Asset Pricing*. Princeton and Oxford: Princeton University Press.

Federal, Reserve. 2014. "3 Month Tresury Bill: Secondary Market Rate." `http://research.stlouisfed.org/fred2/series/TB3MS/`.

Lettau, M. and S. Van Nieuwerburgh. 2008. "Reconciling the Return Predictability Evidence." *The Review of Financial Studies* 21:1607–1652.

Sarkka, S. 2012. "Bayesian Optimal Filtering Equations and Kalman Filter." `http://www.lce.hut.fi/~ssarkka/course_k2012/handout3.pdf`.

Shiller, R. 2014. "Online Data Robert Shiller." `http://www.econ.yale.edu/~shiller/data.htm`.