UNIVERSITY OF COPENHAGEN
DEPARTMENT OF MATHEMATICAL SCIENCES

Master's thesis

Clara Egehoved Tørsløv

Market timing strategies and the business-cycle

Mathematics-Economics

Date: 20. December 2022

Advisors: Rolf Poulsen (UCPH), Laurs Leth (ATP) & Kasper Lund-Jensen (ATP)

Abstract

In this thesis we identify business cycle turning points by applying a state-space model, which incorporates a broad range of growth signals from real-time macroeconomic data of various frequencies. In this way, by analyzing the co-movement of several indicators, we will be able to catch business cycle fluctuations and obtain a good measure of economic activity. Based on the business cycle estimates we obtain two one-month-ahead forecasts of excess return. During NBER recessionary periods the equity premium forecasts sharply drop, mostly to levels negative or near zero, whereafter they sharply rise again to higher than average levels. Strategies exploiting the business cycle estimates and equity premium forecasts produce positive alphas and significantly increase Sharpe ratios over unconditional/passive benchmarks. The strategies profit by successfully timing the market to decrease recessionary losses. We further extend the existing literature by translating the important business cycle algorithm (the Gibbs Sampling approach), originally proposed by Kim (1999) in GAUSS and further developed by Gomez-Cram (2022) in MatLab, to the modern programming language, Python.

Acknowledgements

I am immensely grateful to my supervisors from ATP, Laurs Leth and Kasper Lund-Jensen, who have provided amazing guidance and valuable insights during the project period. Also a big thank you and much appreciation to the whole Liquid Markets team for the warm welcome, and for introducing and including me in the work of the team. I'm very grateful for the opportunity to write this thesis with ATP.

I would also like to thank my supervisor from University of Copenhagen, Rolf Poulsen, for his guidance throughout this thesis and for providing a lot of valuable material and comments, and for always being available when needed.

Thank you to Lasse Heje Pedersen for his good comments and ideas.

A big thank you to Zhuorong Lin, Maja N. Andersen, Sophia Nitsche and Christian Tørsløv for comments and corrections.

Table of content

1 Introduction **5**

2 The business cycle **9**
2.1 Model set up . . . 9
2.2 Recession probability . . . 11
2.3 Posterior Inference . . . 12
2.3.1 Markov Chain Monte Carlo . . . 13
2.4 The Gibbs sampling approach . . . 15
2.4.1 Step 1: Generating the common component . . . 17
2.4.2 Step 2: Generating the specific macro parameters . . . 21
2.4.3 Step 3: Generating the economic regime indicator . . . 23

3 Expected excess return **32**
3.1 Model set up . . . 32
3.1.1 Expected excess returns estimate . . . 33
3.1.2 State-space representation . . . 33
3.2 Posterior Inference . . . 34
3.2.1 The Kalman filter . . . 34
3.2.2 Metropolis-Hasting algorithm . . . 36
3.3 Regression-based forecast . . . 42

4 Timing the market **45**
4.1 Simple Timing . . . 45
4.2 Dynamic Timing . . . 50
4.3 Future research . . . 57
4.3.1 Factor Investing . . . 57
4.4 Data description . . . 61

5 Conclusion **63**

A Appendix **64**

Chapter 1

Introduction

There is a large literature discussing whether active management is beneficial (see for example "Active Portfolio Management and Positive Alphas: Fact or Fantasy?" by Robert Jarrow (2010) Robert A. Jarrow [36]). The goal of active management is to obtain an excess return compared to that of a passive benchmark, by cleverly deviating from such passive benchmark.[1] Grinold (1989) [35] states a fundamental law of active management (ignoring for example transactions costs and trading restrictions),

$$\text{Information Ratio} \approx \text{Information Coefficient} \cdot \sqrt{\text{Breadth}}$$

The information coefficient is the correlation of the portfolio managers forecast with the actual returns. The breadth is the number and frequency of securities traded. Thus the alpha generated by a portfolio manager depends on how well she forecasts and how often and wide she bets. Ang continues a discussion of the reality of the assumptions and the repercussion thereof, ending with the conclusion[2]

The key to generating alpha is forecasting. That requires superior information or superior ability to process public information. - Andrew Ang in Asset Management (2014) [1]

An assumption in the widely used Capital Asset Pricing Model (CAPM) is that information is free and available to all investors. However, the CAPM is derived in a stylized efficient market, and empirically we do not see for example the information efficiency. In reality, data, and especially deriving usable information thereof, is costly and certainly not possible for all investors.[3] In this thesis we acquire, process and analyze several macroeconomic series with the goal to time the market.

For many years the consensus in the literature has been that stock predictability is unstable. Goyal and Welch (2008) [21] study the performance of variables that have been suggested by the academic literature to be good predictors of the equity premium and find that these models would not have helped an investor with access to available information to profitably time the market. Looking back in past literature, forecasting the equity premium takes up great space. See for example the popular paper Fama and French (1988) [7] or, for an overview of several methods, Rapach and Zhou (2013) [4]. The common finding is, that predictability primarily arises around

[1] An example of a passive benchmark could be holding the market portfolio, e.g. the SP 500.
[2] For the full discussion see A. Ang (2014) [1], Asset Management Chapter 10.
[3] A. Ang (2014) [1].

recessions, i.e. that expected equity returns vary over the business cycle. Fama and French (1989) [6] shows that expected returns on stocks and bonds have a clear business-cycle pattern. Campbell and Diebold (2009) [39] find that expected business conditions consistently affect expected excess returns in a counter-cyclical fashion. More interestingly, Lustig and Verdelhan (2012) [16], detect a significant pattern arising around recessions, where realized excess return is negative at the peak of recessions and sharply rises in the following quarters. Moench and Stein (2021) [30] further examine this pattern and shows that the U.S. equity premium follows a pronounced v-shape around the beginning of NBER recessions. This v-shape is clearly seen in Figure 1.1 (Figure 1 of Moench and Stein (2021) [30]), which shows the arithmetic average and median of the log equity premium around the 10 latest recessions (not including 2020), over different time windows. The figure shows a clear v-shaped pattern for the cumulative 6- and 12-month ahead horizons, indicating that the equity premium sharply drops at the onset of recessions however just as strongly recovers afterwards.

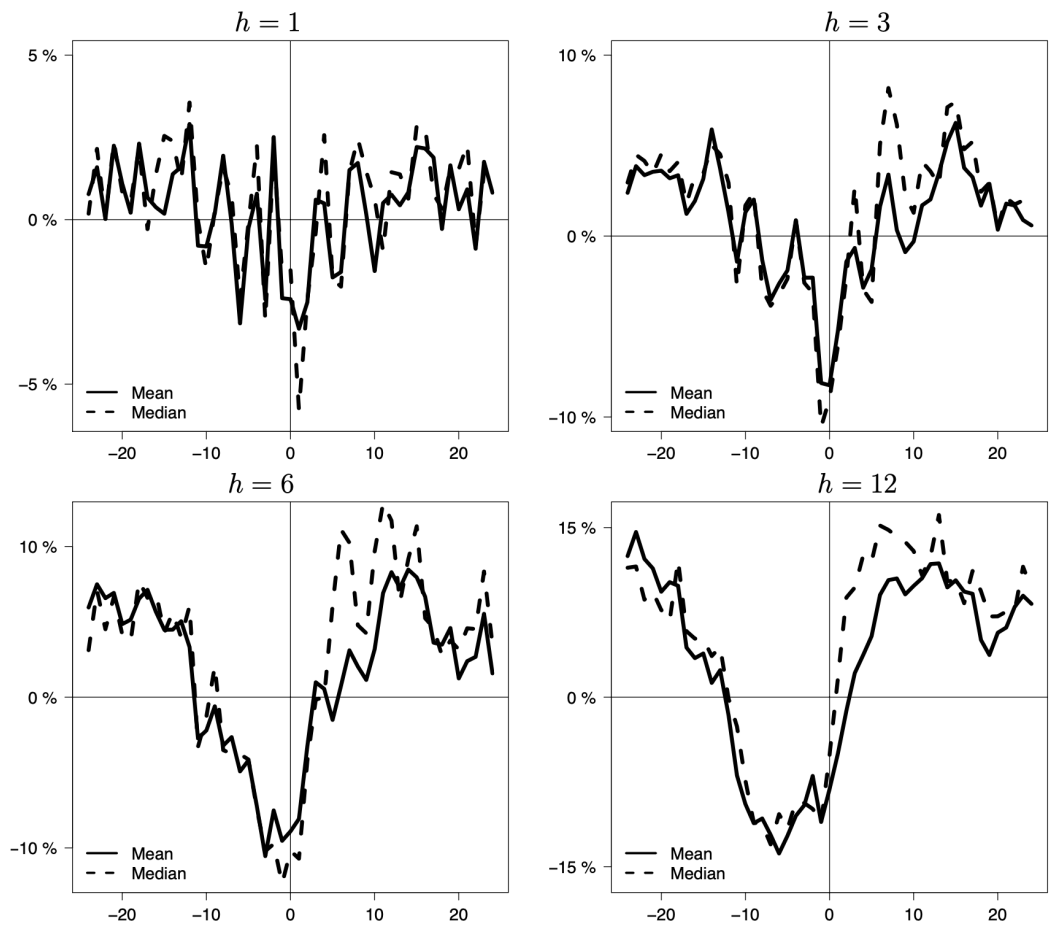


Figure 1.1: Figure and figure text from Moench and Stein (2021) [30]. *"This figure presents the arithmetic average and median of the (cumulative) log equity premium around the 10 recessions in the sample from 1951:3 to 2019:12. The equity premium is the difference between value-weighted returns on the S&P500 index (including dividends) and the Treasury bill rate. The vertical axis depicts* $\sum_{j=0}^{h-1} r_{t+j}$ *for* $t = 24, \ldots, 1, 0, 1, \ldots, 24$*, whereby* $j = 0$ r_{t+j} *is the log equity premium in month* $t + j$*. The horizontal axis displays the 24 months before and after a business cycle peak - with* $t = 0$ *referring to the first month of a NBER-dated recession. Results are shown for* $h = 1, 3, 6, 12$*."*

This evidence strongly suggests that, if we can predict recessions, one should be able to time the market, by selling at the start of recessions and increase risk-taking at a later date. The pattern of Figure 1.1 is found using the true recession dates of the National Bureau of Economic Research (NBER). These dates are often released with significant delay, and is therefore not available for real-time trading. Previously for example the interest rate, industrial production growth or GDP has been used in the literature as an indicator of economic activity (and thereby recession). In this thesis we identify business cycle turning points by applying a state-space model for the business cycle, which incorporates a broad range of growth signals from real-time macroeconomic data of various frequencies. In this way, we argue that by analyzing the co-movement of several indicators, rather than just one, our common growth component catches business cycle fluctuations and thereby is a good measure of economic activity. This common growth component is, amongst other, used to compute a real-time recession probability, seen in Figure 1.2. These real-time probabilities are applied, directly and indirectly, to obtain two equity premium forecasts. We construct timing strategies to analyze the economic relevance of our work. We find that our real-time recession probability carry statistical significant economic relevance (on a 5% level) and, we further produce significant positive CAPM-α on a 1% and 5% level. The strategies benefit by correctly timing the market to reduce risk-taking during bad economic times, thus avoiding stock market crashes.

The business cycle algorithm applied is originally proposed by Kim (1999) [25], in the old GAUSS programming language, and further developed in the MatLab language by Gomez-Cram (2022) [14]. I have further translated it to the modern programming language Python. As concluded in Hilpisch (2016) [46] "Computational Finance Why Python Is Taking Over", Python is becoming more and more important in finance and in particular in computational finance. Hilpisch lists multiple reasons as to why this is the case and concludes that Python is very likely to stay in finance for quite a while. Niknami (2019) [29] compares Python to the C++ programming language, which is also popular in computational finance, and finds that Python solves the problem significantly faster.

The thesis is organized as follows. Chapter 2 introduces the model of the business cycle. This contains, amongst other, introduction and derivation of the two important macro-economic indicators: the common growth component and the real-time filtered probability of recession. Chapter 3 introduces the model of the excess return. This contains, amongst other, introduction and derivation of the two 1-month ahead forecast of excess return. Chapter 4 employs the computed estimates in different trading strategies to test the economic relevance thereof.

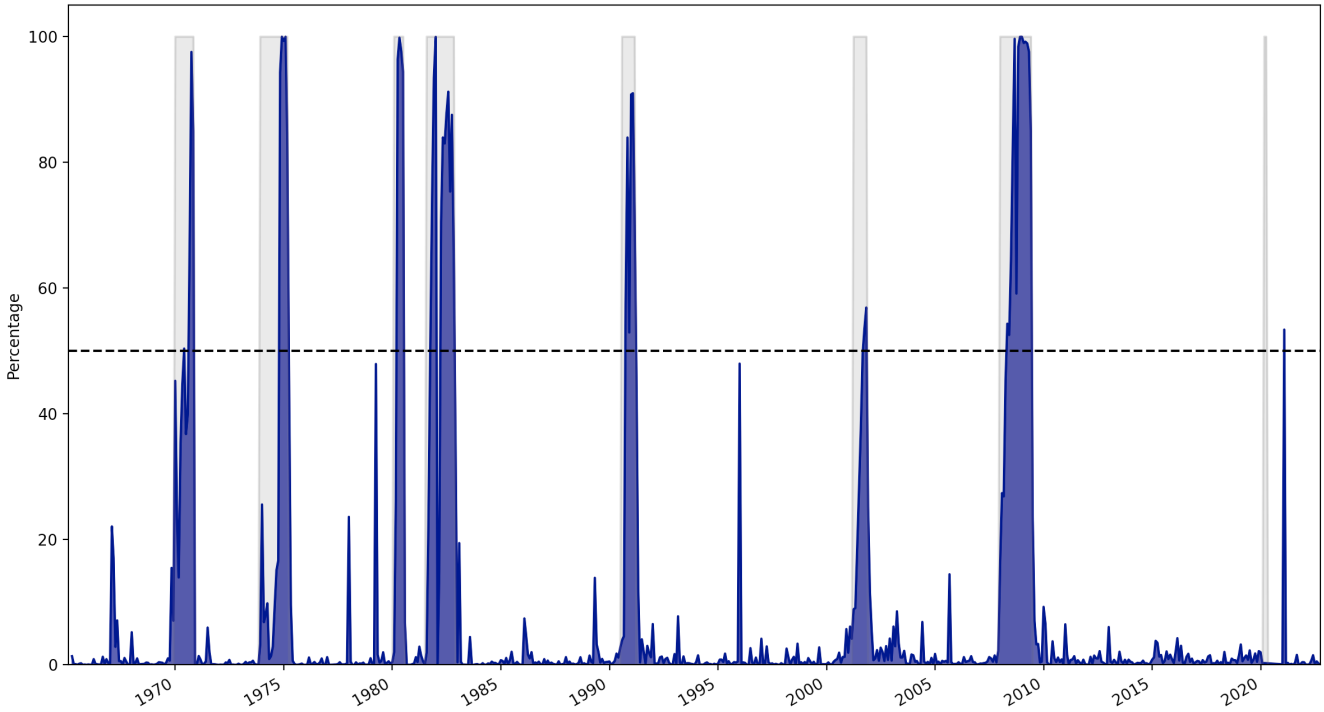


Figure 1.2: Filtered probability of recession. *Plot of the real-time recession probability, as introduced and estimated in Chapter 2. The grey bars indicate NBER recession dates. The sample period runs from January 1965 to September 2022, excluding months March 2020 to November 2020.*

Chapter 2

The business cycle

2.1 Model set up

One of the main outputs of this paper, and a key ingredient in the later trading strategy, is what we will refer to as the common growth component. The growth rate of the common component is highly dependent on whether the economy is in an expansionary or recessionary regime. However, the component is not observable and we therefore need to make inference through what we can observe - the macroeconomic variables.

Let $y_{i,t}$ be the i'th macroeconomic variable at time t. We assume each macro indicator to depend on a non-observable (latent) specific component, $e_{i,t}$, and the latent common growth component, z_t, where the variable, γ_i, captures the vulnerability of the i'th macro indicator to the common component. Then the N macroeconomic variables are assumed to evolve according to

$$y_{i,t} = \gamma_i z_t + e_{i,t} \qquad i = 1, \ldots, N \tag{2.1}$$

The specific components are assumed to follow a first order autoregressive process

$$e_{i,t} = \psi_i e_{i,t-1} + \sigma_{i,t} \cdot \epsilon_{i,t}, \qquad \epsilon_{i,t} \overset{iid}{\sim} \mathcal{N}(0,1) \tag{2.2}$$

We assume the process of the common growth component to depend on a latent economic regime indicator, S_t. This indicator is a Markov process, which signals whether we are in a period with broad economic growth and low volatility (the expansionary regime) or broad economic contractions and high volatility (the recessionary regime). In summary, we assume the common growth component to be generated by a first order autoregressive process with Markov-switching deviations in the conditional mean, persistence, and volatility (to account for regime switching).

$$z_t = \mu_{z,S_t} + \phi_{z,S_t} \cdot (z_{t-1} - \mu_{z,S_{t-1}}) + \sigma_{z,S_t} \cdot \epsilon_{z,t}, \qquad \epsilon_{z,t} \overset{iid}{\sim} \mathcal{N}(0,1) \tag{2.3}$$

We assume the shocks to be independent, $\epsilon_{i,t} \perp\!\!\!\perp \epsilon_{z,t} \, \forall i, t,$. We define the conditional mean and volatility to be in alignment with the regime characteristics and ensure that conditional volatility stays positive in both regimes

$$\mu_{z,S_t} = \mu_{z_0} + \mu_{z_1} \cdot \mathbf{1}_{\{S_t = \text{Recession}\}} \qquad \mu_{z_0} > \mu_{z_0} + \mu_{z_1} \tag{2.4}$$

$$\begin{aligned} \sigma^2_{z,S_t} &= \sigma^2_{z_0}(1 - \mathbf{1}_{\{S_t=\text{Recession}\}}) + \sigma^2_{z_1}\mathbf{1}_{\{S_t=\text{Recession}\}} \\ &= \sigma^2_{z_0} \cdot (1 + h_{\sigma_z} \cdot \mathbf{1}_{\{S_t=\text{Recession}\}}) \qquad h_{\sigma_z} > -1 \end{aligned} \tag{2.5}$$

Lastly, we assume that the transitions between the two regimes progress according to a two-by-two transition matrix consisting of the following transition probabilities

$$\begin{aligned} p &= \mathbb{P}[S_t = \text{Recession} \,|\, S_{t-1} = \text{Recession}] & q &= \mathbb{P}[S_t = \text{Expansion} \,|\, S_{t-1} = \text{Expansion}] \\ 1-p &= \mathbb{P}[S_t = \text{Expansion} \,|\, S_{t-1} = \text{Recession}] & 1-q &= \mathbb{P}[S_t = \text{Recession} \,|\, S_{t-1} = \text{Expansion}] \end{aligned} \tag{2.6}$$

Let Θ_{bc} denote the vector of parameters associated with the business cycle as specified by equations (2.1), (2.2) and (2.3), and $\mathcal{F}_t$ denote the information set available at time t (including the macro variables).

State-space representation

We can represent this model as a state-space model, and in doing so, this opens up a wide toolbox. State-space models are capable of dealing with a very wide range of time series models that are, otherwise, potentially difficult to handle. We assume that the dynamics of a given system is determined by one or more unobservable variables, which in turn are related to an observable time series. Originally, state-space methods were employed by engineers and physicists to infer on the state of noisy systems, fx. to track orbiting satellites' speed and location, by using observable data and the theoretical laws of physics. The methodology of the state-space lies in defining the theoretical relation between the observable and unobservable.

The state-space representation consists of two equations: A measurement equation and a state-transition equation. The first describes the relationship between the observed and unobserved variables. The latter describes the relationship between the unobserved state variables (also referenced as the state vector).

$$\text{(Measurement equation)} \quad y_t = A_t \cdot (H_0 + H_1 \cdot \beta_t + \nu_t), \qquad \nu_t \overset{iid}{\sim} \mathcal{N}(0, R) \tag{2.7}$$

$$\text{(State-transitions equation)} \quad \beta_t = F_{0,S_{t-1},S_t} + F_{1,S_t} \cdot \beta_{t-1} + \omega_t, \qquad \omega_t \overset{iid}{\sim} \mathcal{N}(0, Q_{S_t}) \tag{2.8}$$

The subscript indicates that some model parameters depend upon the latent economic regime indicator, S_t. The measurement equation is often also referred to as the observation equation, as it describes the link between the observed variables/data, y_t, and the unobserved state variable, β_t. It depends on the state vector, β_t, through a design matrix, H_1. A_t is a selection matrix, which accounts for changes in the observations, in our case to account for changes in frequency between monthly and quarterly macroeconomic data. The state-transition equation, which describes the dynamics of the state variable, takes the form of a first order differential equation. F_{1,S_t} is referred to as the transition matrix. The independent white noise sequences, ω_{t+1} and ν_{t+1} are referred to as the innovations (sometimes distinguished by the observation and state innovation).

Above defined business cycle model, (2.1), (2.2) and (2.3), is not identified and therefore cannot uniquely be cast into state-space form.[1] We therefore follow Gómez-Cram (2022) [14] and express the macroeconomic data in deviations from mean, i.e. equation (2.1) simplifies to

$$y^*_{i,t} = y_{i,t} - \psi_i \cdot y_{i,t-1} = \gamma_i \cdot z_t - \gamma_i \cdot \psi_i \cdot z_{t-1} + \sigma_i \cdot \epsilon_{i,t}$$

[1] As the means of the variables are overdetermined in above parameterization, see Kim and Nelson (1998) [24] page 189 for further explanation.

Now the business cycle model can be represented as a state-space model by applying the following matrices

$$\beta_{t+1} = \begin{bmatrix} z_{t+1} & z_t & z_{t-1} & z_{t-2} & z_{t-3} & z_{t-4} & \sigma_1\epsilon_{1,t+1} & \sigma_2\epsilon_{2,t} & \sigma_2\epsilon_{2,t-1} & \sigma_2\epsilon_{2,t-2} & \sigma_2\epsilon_{2,t-3} & \sigma_2\epsilon_{2,t-4} \end{bmatrix}^T_{12\times 1}$$

Given this state vector, the state-space matrices have the following specification

$$H_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}_{2\times 1} \qquad R = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}_{2\times 2}$$

$$H_1 = \begin{bmatrix} \gamma_1 & -\gamma_1\psi_1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{3}\gamma_2 & \frac{2}{3}\gamma_2 - \frac{1}{3}\gamma_2\psi_2 & \gamma_2 - \frac{2}{3}\gamma_2\psi_2 & \frac{2}{3}\gamma_2 - \gamma_2\psi_2 & \frac{1}{3}\gamma_2 - \frac{2}{3}\gamma_2\psi_2 & -\frac{1}{3}\gamma_2\psi_2 & 0 & \frac{1}{3} & \frac{2}{3} & 1 & \frac{2}{3} & \frac{1}{3} \end{bmatrix}_{2\times 12}$$

$$F_{0,S_t,S_{t+1}} = \begin{bmatrix} \mu_{z,S_{t+1}} - \phi_{z,S_{t+1}} \cdot \mu_{z,S_t} \\ \mathcal{O}_{11\times 1} \end{bmatrix}_{12\times 1}$$

$$F_{1,S_{t+1}} = \begin{bmatrix} \phi_{z,S_{t+1}} & 0 & 0 & 0 & 0 & 0 & \\ 1 & 0 & 0 & 0 & 0 & 0 & \\ 0 & 1 & 0 & 0 & 0 & 0 & \\ 0 & 0 & 1 & 0 & 0 & 0 & \mathcal{O}_{6\times 6} \\ 0 & 0 & 0 & 1 & 0 & 0 & \\ 0 & 0 & 0 & 0 & 1 & 0 & \\ & & \mathcal{O}_{6\times 6} & & & & \mathcal{O}_{6\times 6} \end{bmatrix}_{12\times 12}$$

$$Q_{S_{t+1}} = \mathcal{O}_{12\times 12} \quad \text{with} \quad Q_{S_{t+1}}[1,1] = \sigma^2_{z,S_{t+1}}, \quad Q_{S_{t+1}}[7,7] = \sigma_1^2, \quad Q_{S_{t+1}}[8,8] = \sigma_2^2$$

2.2 Recession probability

Another main output of this paper is the filtered probability of being in a recessionary regime at time t, given the knowledge up to time t, $\hat{\pi}_{t|t}$. We will use this quantity as dynamic forecasting weights in a trading strategy later. This probability is highly dependent on the specific business cycle parameters, Θ_{bc}. Given a set of parameters the filtered probability of being in a recessionary regime at time t is given by

$$\hat{\pi}_{t|t}(\Theta_{bc}) = \mathbb{E}[S_t = \text{Recession} \,|\, \Theta_{bc},\, \mathcal{F}_t]$$

Let $p(S_t \,|\, \Theta_{bc}, \mathcal{F}_t)$ denote the posterior density of the regime indicator, S_t. Then we can write out the Bayesian expectation

$$= \int \mathbf{1}_{\{S_t = \text{Recession}\}} p(S_t \,|\, \Theta_{bc},\, \mathcal{F}_t)\, dS_t$$

To remove the dependence on the specific value of business cycle parameters, we can integrate it out, making use of the posterior density of the parameters, $p(\Theta_{bc} \mid \mathcal{F}_t)$. The posterior distribution is sometimes referred to as the Bayesian likelihood function, due to it's properties. The posterior distribution inform us about the likelihood of the set of parameters, given the information set. In a way, we are using the posterior distribution as weights in the integral. This means we assign a higher weight to recession probability estimates, where the set of parameters that led to such estimate, has a high likelihood and reverse. We obtain the posterior mean.

$$\begin{aligned}\hat{\pi}_{t|t} &= \int \hat{\pi}_{t|t}(\Theta_{bc}) p(\Theta_{bc} \mid \mathcal{F}_t)\, d\Theta_{bc} \\ &= \int \int \mathbf{1}_{\{S_t = \text{Recession}\}} p(S_t \mid \Theta_{bc},\, \mathcal{F}_t) p(\Theta_{bc} \mid \mathcal{F}_t)\, d\Theta_{bc}\, dS_t \end{aligned} \tag{2.9}$$

Thus, in order to obtain the estimate we are interested in, we need to do two things: to conduct posterior inference (twice) and then to use the obtained posterior distributions to calculate expectations.

2.3 Posterior Inference

Bayesian models are probabilistic models as they are built on probabilities. The idea is, that probabilities are the mathematically correct tool to model uncertainty. The statistical conceptualization is that all data is generated from some underlying *true* probability distribution, which depend on a set of unobservable parameters. Inference is then the process of estimating the values of those parameters, given a sample from this *true* probability distribution. After conducting Bayesian inference, what we end up with is thus an approximation of the underlying *true* probability distribution, given the sample. The posterior distribution is the output of the Bayesian analysis and is said probability distribution.

A probability distribution is a mathematical object that describes how likely a given event is.[2]

The primary ingredient in posterior inference is Bayes' Theorem, which states

$$\text{Posterior} = \frac{\text{Likelihood} \cdot \text{Prior}}{\text{Marginal likelihood}}$$

$$p(\Theta \mid \tilde{y}_t) = \frac{p(\tilde{y}_t \mid \Theta) \cdot p(\Theta)}{p(\tilde{y}_t)}$$

The marginal likelihood can be seen as a normalization factor. Often Bayesian statisticians do not care about this, as the goal is often to infer on the relative values of the parameters and not the absolute. The same applies in this thesis (unless otherwise stated). Hence, we can express Bayes theorem as a proportionality instead,

$$p(\Theta \mid \tilde{y}_t) \propto p(\tilde{y}_t \mid \Theta) \cdot p(\Theta)$$

Many, if not most, Bayesian models does not result in analytical tractability. To conduct posterior inference, when there is no closed form solution, we employ computational methods, more specifically Markov Chain Monte Carlo methods. These are methods to obtain approximate samples from any target density.

[2] O. Martin (2018) [28]

2.3.1 Markov Chain Monte Carlo

The coverage of Markov Chain Monte Carlo methods alone could be an entire thesis project or even a book. In fact several books have been written on Markov Chains and Monte Carlo methods. Here we will only provide a brief description to provide a basic understanding going forward. Later, we will dig deeper into the actual employment of the algorithm on our specific problem.

A Markov Chain is a stochastic model describing a sequence of possible events in which the probability of each event depends only on the state attained in the previous event.[3]

In mathematical terms this means that, for a sequence of random vectors, $X_1, X_2, X_3, \ldots,$[4][5]

$$\mathbb{P}(X_{t+1} = x \,|\, X_1 = x_1, \ldots X_t = x_t) = \mathbb{P}(X_{t+1} = x \,|\, X_t = x_t)$$

Each vector, X_i, is referred to as a state of the chain and the possible values this vector can attain is referred to as the state space of the chain (not to be confused with the state-space model). This state space could for example be $\mathcal{S} = \{\text{Expansion}, \text{Recession}\}$. A simple example of a Markov chain is the random walk, i.e. $X_{t+1} = X_t + \epsilon_t$, $\epsilon_t \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$, or the price of a stock. The real reason we need to cover Markov Chains, is because the state-space model has exactly those properties. Meaning the state vectors, as defined in the transition equation of the state-space model (2.8), and further the observations, of the measurement equation (2.7), both define a Markov chain.

Associated with the Markov chain is also the transitioning between states. This is defined by a transition probability density function, $T(X_{t+1} \,|\, X_t)$, and sometimes represented by a transition matrix, whose $ij-th$ element is $p_{ij} = \mathbb{P}(X_{t+1} = i \,|\, X_t = j)$. This could for example be transitioning from Expansion to Recession. If the transition probability, p_{ij}, does not depend on t we say the chain is homogeneous, however this is usually not the case. The most important use of the transition density (sometimes also referred to as the density kernel) is convergence of the chain. The choice of density kernel is important as this ensures that the limit distribution of the chain is the target density. A derivation and discussion of the convergence can be found many places, see fx. the chapters on MCMC in N. Heard (2022) [17] or K. Triantafyllopoulos (2021) [43]. See especially Definition 5.7 and Proposition 5.2 in N. Heard (2022) [17], which in summary states that if the density kernel satisfies the detailed balance condition, then the stationary distribution of the chain is the target distribution. For a proof of convergence, i.e. that the Markov chain converges to the stationary distribution, see Theorem 4.9 in D. Levin et al. (Markov Chains and Mixing Times, 2008) [26].

The keyword here is limit, meaning we must have a *reasonable* amount of samples before we can say the chain have converged. We therefore discard the initial samples, from now on referred to as the burn-in samples or period, as these are not assumed to follow the target distribution. The remaining Markov Chain samples are saved and used in the Monte Carlo estimation. See Definition 5.1 *(Monte Carlo estimate of an expectation)* in Heard (2022) [17].

Monte Carlo methods are approximate, sampling-based approaches for evaluating expectations.[6]

[3] For further explanation on Markov Chains see Wikipedia [45]

[4] Assuming the conditional probabilities are well defined.

[5] An example of such sequence could be the time series or the state process, dynamics in equation (2.7) and (2.8).

[6] Section 5.2 in Heard (2022) [17]

So, long story short, we use the Markov Chain methodology to generate samples from the target distribution and the Monte Carlo methodology to compute an approximation of the integral by averaging over the generated samples.[7] Throughout this thesis we will employ the two most common algorithms for constructing a suitable Markov Chain - the first being Gibbs sampling and the second is the Metropolis-Hastings algorithm. The latter will be introduced and applied in section 3.2.2.

Gibbs sampling

By Proposition 5.3 in N. Heard (2022) [17], Gibbs Sampling satisfies the detailed balance condition and thus samples from the Gibbs Sampling Markov chain will converge to the target distribution. An assumption of the Gibbs sampler is that we can sample from the conditional distributions. The Gibbs Sampler, as described in K. Triantafyllopoulos (2021) [43] on page 243,

1. Initialize the iteration counter, $i = 0$, with initial sample $X^{(0)} = [X_1^{(0)}, \ldots, X_k^{(0)}]$.

2. Obtain a sequence of vectors $X^{(i)}$ conditional on $X^{(i-1)}$, by sampling one value at a time, $X_j^{(i)}, j = 1, \ldots, k$, according to

$$\begin{aligned}
&X_1^{(i)} \sim p(X_1 \,|\, X_2^{(i-1)}, X_3^{(i-1)}, \ldots, X_k^{(i-1)}) \\
&X_2^{(i)} \sim p(X_2 \,|\, X_1^{(i)}, X_3^{(i-1)}, \ldots, X_k^{(i-1)}) \\
&\vdots \\
&X_k^{(i)} \sim p(X_k \,|\, X_1^{(i)}, X_2^{(i)}, \ldots, X_{k-1}^{(i)})
\end{aligned}$$

3. Set the counter $i = i + 1$ and go to step 2 until convergence.

Each generated vector of vectors, $X^{(i)}$, is a draw from a Markov chain. By employing this algorithm one would obtain '$M-$burn-in' (if i runs from 1 to M) approximate samples from the target distribution.

One thing that we need to beware of is the dependency between samples. There are several methods that we either need to be careful with or even cannot use. An example is the Monte Carlo standard error, as this assumes independent samples and thus the results would not be valid.[8] Further, the first several samples are dependent on the choice of initial value, $X^{(0)}$, which is also one of the reasons we discard the burn-in samples. This means that the initial value is not of significant importance for the result, however it is in regards to the speed of convergence.

A clear advantage of the dependent samples is the convergence of the algorithm. Assume we have sampled from a high density area and accepted the candidate. We can then use this information to again sample from this high density area, whereas if the samples were independent

[7] We assume the reader knows Monte Carlo methodology and the argumentation belonging to. As a Math-Econ student Monte Carlo methods, and especially the law of large numbers, is lodged into my brain for eternity. We probably need to mention it here anyway, so as $n \to \infty$ then $\frac{1}{n}\sum_{i=1}^{n} x_i \to \mathbb{E}[X]$.

[8] The Monte Carlo standard error can be found in N. Heard (2022) [17] Definition 5.2.

we would have started from scratch every single time and thus rejected a very large amount of samples. By utilizing the information in the previous sample, the algorithm will be much more efficient. There are several ways to circumvent the dependency structure - one example is the burn-in period, and another could be to only adopt every m accepted candidate. Further discussion can be found in e.g. M. Sugiyama (2016) [41].

As mentioned in N. Heard (2022) [17], Gibbs sampling should be used with caution. We therefore start by checking if we can replicate the results in Gómez-Cram (2022) [14] before continuing with our analysis. This comparison can be seen in **??**, where we find that they strongly coincide. This leads us to the specification of the algorithm driving this thesis.

2.4 The Gibbs sampling approach

Consider the business cycle model as described in section 2.1. Both the parameters of the model, the latent Markov-Switching variable and the state variable are treated as random variables. To obtain Bayesian inference on these random variates, we need the joint distribution. As this is tricky we employ the Gibbs Sampling approach. This tool is used when sampling from the joint distribution is impossible (or close to impossible) but sampling from a conditional distribution is feasible. The algorithm is introduced by Kim and Nelson (1999) [25] and works in the following way. [9]

Apply the following three steps until convergence occurs, with arbitrary starting values for the model parameters.[10] Let the tilde operator denote a vector containing values of the given variable up to the time of the subscript, fx. $\tilde{z}_T = [z_1, \ldots, z_T]$.

Step 1. Generate the common component, $\tilde{z}_T$, conditional on all parameters of the model, the economic regime indicator $\tilde{S}_T$ and the macroeconomic data $\tilde{y}_{i,T}$ for $i = 1, \ldots, N$.

Step 2. For $i = 1, \ldots, N$: Generate the parameters associated with the macroeconomic equations (2.1) and (2.2), conditional on the common component, $\tilde{z}_T$, and data on the i'th macroeconomic variable, $\tilde{y}_{i,T}$.

Step 3. Generate the economic regime indicator, $\tilde{S}_T$, the parameters associated with the common component, equations (2.4) and (2.5), and the transition probabilities, equation (2.6), conditional on the common component, $\tilde{z}_T$.

Each step will be further explained when relevant. As we will see, the conditioning features of each step allows us to employ much simpler tools for computation of each block of variables.

[9] See Kim and Nelson (1999) [25] Chapter 10 page 242.

[10] We discard the first 15.000 samples, and thus the starting values of the model parameters are not of significant importance. This is also mentioned in Kim and Nelson (1999) [25]

Algorithm flowchart

The algorithm described in this section can be summarized in the following flowchart.

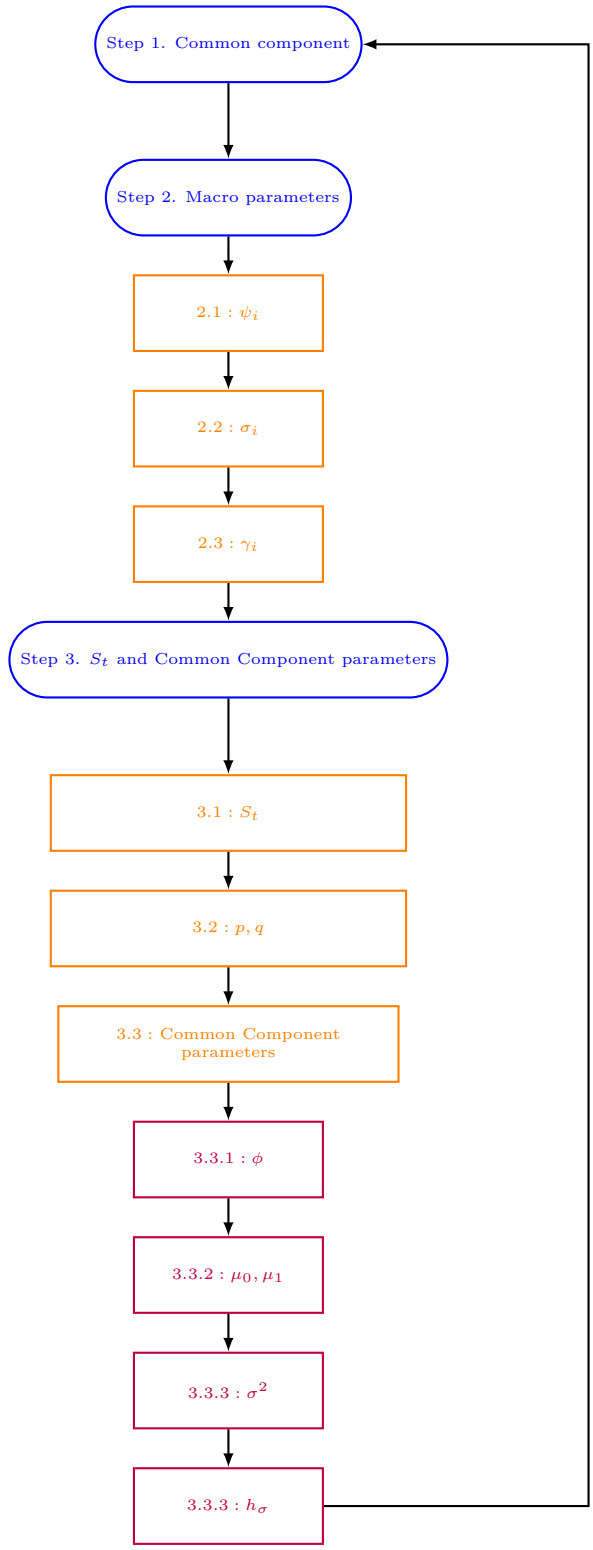


Figure 2.1: Algorithm flowchart

2.4.1 Step 1: Generating the common component

If we consider the state-space model given by equations (2.7) and (2.8), some model parameters depend on the latent regime indiator, S_t. As we now condition on this variable, above model reduces to a Gaussian linear state-space model with a known dummy, S_t, to account for structural breaks.

$$y_t = A \cdot (H_0 + H_1 \cdot \beta_t + e_t), \qquad e_t \overset{iid}{\sim} \mathcal{N}(0, R) \tag{2.10}$$

$$\beta_t = F_0 + F_1 \cdot \beta_{t-1} + v_t, \qquad v_t \overset{iid}{\sim} \mathcal{N}(0, Q) \tag{2.11}$$

We note how, clearly, the observations, y_t, depend linearly on the state variable, β_t, and further that the state variable dynamic is linear too. Further note that the innovations are Gaussian. We therefore have a Gaussian linear state-space model, for which we have a toolbox. The first element of the state vector, β_t, is the common component, z_t. Thus, if we generate the state vector, $\tilde{\beta}_T$, we generate the common component. As done in Kim and Nelson (1999) [25] we wish to apply the multimove Gibbs Sampler to generate the whole vector $\tilde{\beta}_T$ from their joint distribution. Given the data, $\tilde{y}_T = [\tilde{y}_{1,T}, \ldots, \tilde{y}_{N,T}]$ and the prior distribution of β_0, the joint distribution of $\tilde{\beta}_T$ is given by

$$p(\tilde{\beta}_T \,|\, \tilde{y}_T) = p(\beta_1, \ldots, \beta_T \,|\, \tilde{y}_T)$$

Now, using the chain rule of probability, consecutively, on the joint densities

$$\begin{aligned}
&= p(\beta_T \,|\, \tilde{y}_T) p(\beta_1, \ldots, \beta_{T-1} \,|\, \beta_T, \tilde{y}_T) \\
&= p(\beta_T \,|\, \tilde{y}_T), p(\beta_{T-1} \,|\, \beta_T, \tilde{y}_T) p(\beta_1, \ldots, \beta_{T-2} \,|\, \beta_{T-1}, \beta_T, \tilde{y}_T) \\
&\vdots \\
&= p(\beta_T \,|\, \tilde{y}_T) p(\beta_{T-1} \,|\, \beta_T, \tilde{y}_T) p(\beta_{T-2} \,|\, \beta_{T-1}, \beta_T, \tilde{y}_T) \ldots p(\beta_1 \,|\, \beta_2, \ldots, \beta_T, \tilde{y}_T)
\end{aligned}$$

Next, by the Markov property of the state variable, β_t, we can deduce that conditional on β_{t-1}, β_{t-2} contains no information beyond that in β_{t-1}. The same argument can be applied to $\tilde{y}_{t-2}$.

$$\begin{aligned}
&= p(\beta_T \,|\, \tilde{y}_T) p(\beta_{T-1} \,|\, \beta_T, \tilde{y}_{T-1}) p(\beta_{T-2} \,|\, \beta_{T-1}, \tilde{y}_{T-2}) \ldots p(\beta_1 \,|\, \beta_2, \tilde{y}_1) \\
&= p(\beta_T \,|\, \tilde{y}_T) \prod_{t=1}^{T-1} p(\beta_t \,|\, \beta_{t+1}, \tilde{y}_t)
\end{aligned}$$

Define the conditional mean and variance of the state variable

$$\beta_{t|t} = \mathbb{E}[\beta_t \,|\, \tilde{y}_T] \tag{2.12}$$

$$P_{t|t} = \mathbb{V}[\beta_t \,|\, \tilde{y}_T] \tag{2.13}$$

As the model is linear and Gaussian it follows from Lemma 1 (which we will cover below in the derivation of the Kalman equations) that the conditional distributions are also Gaussian.

$$\beta_T \,|\, \tilde{y}_T \sim \mathcal{N}(\beta_{T|T}, P_{T|T}) \tag{2.14}$$

$$\beta_t \,|\, \beta_{t+1}, \tilde{y}_t \sim \mathcal{N}(\beta_{t|t,\beta_{t+1}}, P_{t|t,\beta_{t+1}}) \tag{2.15}$$

We can therefore apply the Gaussian Kalman Filter to obtain $p(\beta_T \,|\, \tilde{y}_T)$ and $p(\beta_t \,|\, \beta_{t+1}, \tilde{y}_t)$ to obtain the joint distribution of $\tilde{\beta}_T$ and generate the common component. The Kalman filter is a linear estimation method for generating estimates of a non-observable state vector. The procedure, as described in Kim and Nelson (1999) [25], is

- For $t = 1, \ldots, T$: Employ the Gaussian Kalman Filter to generate $\beta_{t|t}$ and $P_{t|t}$ and save the values. The last iteration gives us $\beta_{T|T}$ and $P_{T|T}$.
- For $t = T$: Generate the last value of the state vector and the common component, remembering that the common component is the first element of the state vector, from

$$\begin{aligned} \beta_T \,|\, \tilde{y}_T &\sim \mathcal{N}(\beta_{T|T}, P_{T|T}) \\ z_T \,|\, \tilde{y}_T &\sim \mathcal{N}(\beta_{T|T}[1], P_{T|T}[1,1]) \end{aligned}$$

- For $t = T-1, \ldots, 1$: Generate the state vector, β_t, and the common component from $(\beta_t \,|\, \beta_{t+1}, \tilde{y}_t)$, using

$$\begin{aligned} \beta_t \,|\, \beta_{t+1}, \tilde{y}_t &\sim \mathcal{N}(\beta_{t|t,\beta_{t+1}}, P_{t|t,\beta_{t+1}}) \\ z_t \,|\, \beta_{t+1}, \tilde{y}_t &\sim \mathcal{N}(\beta_{t|t,\beta_{t+1}}[1], P_{t|t,\beta_{t+1}}[1,1]) \end{aligned}$$

The Gaussian Kalman Filter

The Kalman filter is a widely used recursive algorithm, applied for estimation of non-observable variables. In our case we wish to use it to infer on the latent state variable, β_t, and thereby obtain a predictor of the value of the next observation, y_t. The algorithm works in two steps, prediction and update. Both steps are performed at time t, for $t = 1, \ldots, T$, however which step we are in depends on the information set, meaning whether we have observed the data point, y_t, yet. The observation is assumed to be made at the end of time t.[11] Before initialization of the algorithm the initial values are calculated. The derivation of the Kalman equations and initial values will be covered in separate following sections.

<u>Prediction</u>: We have not yet received the information of y_t, but wish to infer on y_t given the available information. We call this predictor $y_{t|t-1}$.

$$\beta_{t|t-1} = F_0 + F_1 \cdot \beta_{t-1|t-1} \tag{2.16}$$

$$P_{t|t-1} = F_1 \cdot P_{t-1|t-1} \cdot F_1^T + Q \tag{2.17}$$

$$y_{t|t-1} = A \cdot (H_0 + H_1 \cdot \beta_{t|t-1}) \tag{2.18}$$

The new observation, y_t, is made at the end of time t.

$$\eta_{t|t-1} = y_t - y_{t|t-1} \tag{2.19}$$

$$f_{t|t-1} = (A \cdot H_1) \cdot P_{t|t-1} \cdot (A \cdot H_1)^T + R \tag{2.20}$$

<u>Updating:</u> The prediction error, $\eta_{t|t-1}$, is used to update the inference on β_t, to obtain $\beta_{t|t}$. The weight given to this new information is determined by the Kalman gain.

$$\beta_{t|t} = \beta_{t|t-1} + P_{t|t-1} \cdot (A \cdot H_1)^T \cdot f_{t|t-1}^{-1} \cdot \eta_{t|t-1} \tag{2.21}$$

$$P_{t|t} = P_{t|t-1} + P_{t|t-1} \cdot (A \cdot H_1)^T \cdot f_{t|t-1}^{-1} \cdot (A \cdot H_1) \cdot P_{t|t-1} \tag{2.22}$$

$$\text{(Kalman Gain)} \qquad K_t = P_{t|t-1} \cdot (A \cdot H_1)^T \cdot f_{t|t-1}^{-1} \tag{2.23}$$

[11] Durbin and Koopman (2012) [5] chapter 4.3 and Kim and Nelson (1999) [25] Chapter 3.1.2

The Kalman Gain

One observation to make is where uncertainty arises from. This can be inferred from equations (2.17) and (2.20). It's clearly seen how $\beta_{t|t-1}$ is dependent on $\beta_{t-1|t-1}$. Thus, uncertainty connected to $\beta_{t|t-1}$ will be dependent on $P_{t-1|t-1}$, the uncertainty underlying $\beta_{t-1|t-1}$, as well as on Q, the covariance matrix of shocks to β_t. This is clearly seen in equation (2.17) (where $P_{t|t-1}$ depends on $P_{t-1|t-1}$ and Q). Similar reflections can be made in reference to equation (2.20). Here it is clearly seen how uncertainty arises from two locations: the uncertainty underlying $\beta_{t-1|t-1}$, $P_{t-1|t-1}$, and the uncertainty associated with a random shock to y_t, R. The Kalman gain includes both (2.17) and (2.20). Examining the Kalman gain further we find that the weight assigned to the new information, depends positively on the uncertainty surrounding $\beta_{t|t-1}$, $P_{t|t-1}$ and negatively on the uncertainty surrounding shocks to y_t, R.[12] Put in intuitively terms, if the uncertainty of the estimation of $\beta_{t|t-1}$ is high, a larger weight will be put on the new information $\eta_{t|t-1}$. Similarly, if the uncertainty of the new information increases then a smaller weight will be put on this information.

Derivation of the Kalman equations

The key element in the derivation of the Gaussian Kalman filter is the following Lemma 1. A proof of Lemma 1 can be found in Durbin & Koopman (2012) [5] page 78.

Lemma 1. *Assume x and y are jointly normally distributed random vectors with*

$$\mathbb{E}\left[\begin{pmatrix} x \\ y \end{pmatrix}\right] = \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix} \qquad \mathbb{V}\left[\begin{pmatrix} x \\ y \end{pmatrix}\right] = \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{xy}^T & \Sigma_{yy} \end{bmatrix}$$

where Σ_{yy} is a nonsingular matrix. Then the conditional distribution of x given y is normal with mean vector and variance matrix given by

$$\mathbb{E}\left[x|y\right] = \mu_x + \Sigma_{xy}\Sigma_{yy}^{-1}(y - \mu_y) \tag{2.24}$$

$$\mathbb{V}\left[x|y\right] = \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{xy}^T \tag{2.25}$$

Note that we assume non singularity of the covariance matrix, meaning the Q matrix of equation (2.8), in above specified state-space model, must be positive-definite.

Consider the Gaussian linear model given by equations (2.10) and (2.11). Our goal is to determine the conditional mean and variance of β_t. We define the prediction error, also referred to as the one-step ahead forecast error of y_t given $\tilde{y}_{t-1}$. As $e_t \perp\!\!\!\perp \tilde{y}_{t-1}$ it follows

$$\begin{aligned} \eta_{t|t-1} &= y_t - y_{t|t-1} = y_t - \mathbb{E}[y_t \,|\, \tilde{y}_{t-1}] \\ &= y_t - \mathbb{E}[A \cdot (H_0 + H_1 \cdot \beta_t + \underbrace{e_t}_{\mathbb{E}[e_t]=0}) \,|\, \tilde{y}_{t-1}] = y_t - A \cdot (H_0 + H_1 \cdot \beta_{t|t-1}) \end{aligned}$$

[12] Assume for computational simplicity that we are in the one dimensional case. Then $K_t = \frac{P_{t|t-1}(A \cdot H_1)}{P_{t|t-1} \cdot (A \cdot H_1)^2 + R}$ and thus $\frac{\partial K_t}{\partial P_{t|t-1}} = \frac{(A \cdot H_1)R}{(P_{t|t-1} \cdot (A \cdot H_1)^2 + R)^2} > 0$ and $\frac{\partial K_t}{\partial R} = -\frac{(A \cdot H_1)P_{t|t-1}}{(P_{t|t-1} \cdot (A \cdot H_1)^2 + R)^2} < 0$

Note that when $\tilde{y}_{t-1}$ and $\eta_{t|t-1}$ are known, so is $\tilde{y}_t$ and reversely. From this, the following observations results

$$\begin{aligned}
\beta_{t|t} &= \mathbb{E}[\beta_t, \,|\, \tilde{y}_t] = \mathbb{E}[\beta_t, \,|\, \tilde{y}_{t-1}, \eta_{t|t-1}] \\
\mathbb{E}[\eta_{t|t-1}, \,|\, \tilde{y}_{t-1}] &= \mathbb{E}[y_t - A \cdot (H_0 + H_1 \cdot \beta_{t|t-1}), \,|\, \tilde{y}_{t-1}] \\
&= A \cdot (H_0 + H_1 \cdot \beta_{t|t-1}) - A \cdot (H_0 + H_1 \cdot \beta_{t|t-1}) = 0
\end{aligned}$$

By the tower property it follows

$$\mathbb{E}[\eta_{t|t-1}] = A \cdot (H_0 + H_1 \cdot \underbrace{\mathbb{E}[\beta_t]}_{=\mathbb{E}[\mathbb{E}[\beta_t \,|\, \tilde{y}_{t-1}]]}) - A \cdot (H_0 + H_1 \cdot \mathbb{E}[\mathbb{E}[\beta_t, \,|\, \tilde{y}_{t-1}]]) = 0$$

From this it further follows, for $j = 1, \ldots, t-1$

$$Cov(y_j, \eta_{t|t-1}) = \mathbb{E}[y_j \eta_{t|t-1}] - \mathbb{E}[y_j]\mathbb{E}[\eta_{t|t-1}] = \mathbb{E}[\mathbb{E}[y_j \eta_{t|t-1} \,|\, \tilde{y}_{t-1}]] = \mathbb{E}[y_j \mathbb{E}[\eta_{t|t-1} \,|\, \tilde{y}_{t-1}]] = 0$$

We are now ready to apply Lemma 1, with $x = \beta_t$ and $y = \eta_{t|t-1}$, to obtain the mean vector and variance matrix, of the conditional distribution of β_t given $\eta_{t|t-1}$ (and $\tilde{y}_{t-1}$).

$$\begin{aligned}
\beta_{t|t} &= \underbrace{\mathbb{E}[\beta_t \,|\, \tilde{y}_{t-1}]}_{=\beta_{t|t-1}} + Cov(\beta_t, \eta_{t|t-1} \,|\, \tilde{y}_{t-1}) \cdot \mathbb{V}[\eta_{t|t-1} \,|\, \tilde{y}_{t-1}]^{-1} \cdot \eta_{t|t-1} \\
Cov(\beta_t, \eta_{t|t-1}) &= \mathbb{E}[\beta_t \eta_{t|t-1}^T \,|\, \tilde{y}_{t-1}] - \mathbb{E}[\beta_t \,|\, \tilde{y}_{t-1}]\mathbb{E}[\eta_{t|t-1} \,|\, \tilde{y}_{t-1}] \\
&= \mathbb{E}[\beta_t (y_t - A \cdot (H_0 + H_1 \cdot \beta_{t|t-1}))^T \,|\, \tilde{y}_{t-1}] \\
&= \mathbb{E}[\beta_t (A \cdot (H_0 + H_1 \cdot \beta_t) - A \cdot (H_0 + H_1 \cdot \beta_{t|t-1}))^T \,|\, \tilde{y}_{t-1}] \\
&= \mathbb{E}[\beta_t (A \cdot H_1 \cdot \beta_t - A \cdot H_1 \cdot \beta_{t|t-1})^T \,|\, \tilde{y}_{t-1}] \\
&= \mathbb{E}[\beta_t (\beta_t - \cdot \beta_{t|t-1})^T \,|\, \tilde{y}_{t-1}] \cdot (A \cdot H_1)^T \\
&= \underbrace{\mathbb{V}[\beta_t \,|\, \tilde{y}_{t-1}]}_{:=P_{t|t-1}} \cdot (A \cdot H_1)^T
\end{aligned}$$

Next, consider the conditional variance of the prediction error

$$\begin{aligned}
f_{t|t-1} &:= \mathbb{V}[\eta_{t|t-1} \,|\, \tilde{y}_{t-1}] \\
&= \mathbb{V}[A \cdot (H_0 + H_1 \cdot \beta_t + e_t) - A \cdot (H_0 + H_1 \cdot \beta_{t|t-1}) \,|\, \tilde{y}_{t-1}] \\
&= \mathbb{V}[A \cdot H_1 \cdot \beta_t + A \cdot e_t - A \cdot H_1 \cdot \beta_{t|t-1} \,|\, \tilde{y}_{t-1}]
\end{aligned}$$

Remembering that $e_t \sim \mathcal{N}(0, R)$

$$= A \cdot R \cdot A^T + \mathbb{V}[A \cdot H_1 \cdot \beta_t - A \cdot H_1 \cdot \beta_{t|t-1} \,|\, \tilde{y}_{t-1}]$$

Next, using simple variance calculational rules and the fact that $\beta_{t|t-1}$ is considered a constant[13]

$$= A \cdot R \cdot A^T + (A \cdot H_1) \cdot \mathbb{V}[\beta_t \,|\, \tilde{y}_{t-1}] \cdot (A \cdot H_1)^T$$

Collecting this all together we find exactly equation (2.21)

$$\beta_{t|t} = \beta_{t|t-1} + P_{t|t-1} \cdot (A \cdot H_1)^T \cdot f_{t|t-1}^{-1} \cdot \eta_{t|t-1}$$

By Lemma 1, we further get an expression for the conditional variance, equation (2.22)

$$\begin{aligned}
P_{t|t} &= \mathbb{V}[\beta_t \mid \tilde{y}_t] \\
&= \mathbb{V}[\beta_t \mid \tilde{y}_{t-1}, \eta_{t|t-1}] \\
&= \mathbb{V}[\beta_t \mid \tilde{y}_{t-1}] - Cov(\beta_t, \eta_{t|t-1} \mid \tilde{y}_{t-1}) \cdot \mathbb{V}[\eta_{t|t-1} \mid \tilde{y}_{t-1}]^{-1} \cdot Cov(\beta_t, \eta_{t|t-1} \mid \tilde{y}_{t-1})^T \\
&= P_{t|t-1} - P_{t|t-1} \cdot (A \cdot H_1)^T \cdot f_{t|t-1}^{-1} \cdot (A \cdot H_1) \cdot P_{t|t-1}
\end{aligned}$$

Initial values for the Kalman filter
The initial values are not of significant importance, due to the burn-in period. We follow Gómez-Cram (2022) [14] and use the unconditional mean and covariance matrix as initial values. These are calculated prior to initializtion of the algorithm.

2.4.2 Step 2: Generating the specific macro parameters

We consider the business cycle model as described in Section 2.1. If we condition this model on the common component, $\tilde{z}_T$ and the data $\tilde{y}_T$, it collapses into N independent regression equations, each with autocorrelated disturbances. We restate for ease of readability. For $i = 1, \ldots, N$, the model we're considering is

$$\begin{aligned}
y_{i,t} &= \gamma_i \cdot z_t + e_{i,t} \\
e_{i,t} &= \psi_i \cdot e_{i,t-1} + \sigma_{i,t} \cdot \epsilon_{i,t}, \quad \epsilon_{i,t} \overset{iid}{\sim} \mathcal{N}(0,1)
\end{aligned}$$

The idea is then, conditional on the common component and the macroeconomic data, for $i = 1, \ldots, N$,[14]

1. Generate ψ_i conditional on $\sigma_{i,t}$ and γ_i.
2. Generate $\sigma_{i,t}$ conditional on ψ_i and γ_i.
3. Generate γ_i conditional on $\sigma_{i,t}$ and ψ_i.

We start by defining the priors of the macro parameters. When choosing a prior we have to make sure that the prior distribution covers all possible values and that it does not allow any values that are impossible. The latter is very relevant in relation to variance. We follow Gómez-Cram (2022) [14] and assume normal priors for ψ_i and γ_i and for $\sigma_{i,t}$ we will assume an inverse gamma distribution as prior. We argue vaguely that due to the Central Limit Theorem we choose normal priors, but mostly it is because Gaussians are easy to work with, they are a conjugate prior for the Gaussian distribution itself, and, as stated in Martin et al. (2022) [33] it is the largest entropy under the given constraints. For a thorough examination on priors in Bayesian Deep Learning see Vincent Fortuin (2022) [8]. This could be an improvement point for further examination beyond the scope of this thesis. As per the choice of the inverse gamma distribution as a prior for standard deviation, this is what's done in similar articles, see fx Schorfheide (2005) [37] Table 1 on page 405.[15]

[13] $\mathbb{V}(a + x) = \mathbb{V}(x)$ and $\mathbb{V}(a \cdot x) = a^2 \cdot \mathbb{V}(x)$
[14] Kim (1999) [25], section 8.3.3
[15] For an informal forum discussion as to why, the reader is referred to this StackExchange post [40]

For $i = 1, \ldots, N$ we define the following priors, where the parameters are assumed to be known

$$\begin{aligned}
\psi_i \,|\, \sigma_{i,t}, \gamma_i &\sim \mathcal{N}(a_i, A_i) \\
\gamma_i \,|\, \sigma_{i,t}, \psi_i &\sim \mathcal{N}(b_i, B_i) \\
\sigma_{i,t} \,|\, \psi_i, \gamma_i &\sim IG\left(\frac{c_i}{2}, \frac{C_i}{2}\right)
\end{aligned}$$

Our goal is to derive analytical expressions for the posterior distributions, such that we can easily draw samples from these. This is possible as we have defined conjugate priors.[16]

Step 2.1: Generate $\psi_\mathbf{i}$

We combine the two model equations to obtain

$$\begin{aligned}
y_{i,t} = \gamma_i \cdot z_t + e_{i,t} \quad &\Leftrightarrow \quad e_{i,t} = y_{i,t} - \gamma_i \cdot z_t \\
e_{i,t} = \psi_i \cdot e_{i,t-1} + \epsilon_{i,t} \quad &\Leftrightarrow \quad y_{i,t} - \gamma_i \cdot z_t = \psi_i \cdot (y_{i,t-1} - \gamma_i \cdot z_{t-1}) + \epsilon_{i,t}, \quad \epsilon_{i,t} \overset{iid}{\sim} \mathcal{N}(0, \sigma^2_{i,t})
\end{aligned} \tag{2.26}$$

In matrix notation this yields, where I_{T-2} denotes the (T-2)-dimensional identity matrix and T is the sample size,

$$E^* = \psi_i \cdot E^*_{shifted} + \epsilon_i, \quad \epsilon_i \overset{iid}{\sim} \mathcal{N}(0, \sigma^2_{i,t} \cdot I_{T-2}) \tag{2.27}$$

From this (and the prior) the posterior distribution follows

$$\psi_i \,|\, \sigma_{i,t}, \gamma_i, \tilde{y}_T, \tilde{z}_T \sim \mathcal{N}(\bar{a}_i, \bar{A}_i)$$

$$\begin{aligned}
\bar{a}_i &= \left(A_i^{-1} + \sigma_i^{-2} {E^*_{shifted}}^T E^*_{shifted}\right)^{-1} \left(A_i^{-1} a_i + \sigma_i^{-2} {E^*_{shifted}}^T E^*\right) \\
\bar{A}_i &= \left(A_i^{-1} + \sigma_i^{-2} {E^*_{shifted}}^T E^*_{shifted}\right)^{-1}
\end{aligned}$$

Then, generating ψ_i amounts to drawing from this posterior distribution.

Step 2.2: Generate $\sigma_{\mathbf{i,t}}$

From equation (2.27) derived above, combined with the prior, the posterior distribution follows

$$\sigma_{i,t} \,|\, \psi_i, \gamma_i, \tilde{y}_T, \tilde{z}_T \sim IG\left(\frac{\bar{c}_i}{2}, \frac{\bar{C}_i}{2}\right)$$

$$\begin{aligned}
\bar{c}_i &= c_i + (T-2) \\
\bar{C}_i &= C_i + (E^* - \psi_i \cdot E^*_{shifted})^T (E^* - \psi_i \cdot E^*_{shifted})
\end{aligned}$$

Then, generating $\sigma_{i,t}$ amounts to drawing from this posterior distribution.

[16] A conjugate prior is a prior distribution chosen such that the posterior distribution belongs to the same parametric/exponential family.

Step 2.3: Generate γ_i

We consider the model equation

$$y_{i,t} = \gamma_i \cdot z_t + e_{i,t}$$

We apply the first order lag operator to both sides of the equality sign to obtain

$$y_{i,t} - \psi_i \cdot y_{i,t-1} = \gamma_i \cdot (z_t - \psi_i \cdot z_{t-1}) + \epsilon_{i,t}, \quad \epsilon_{i,t} \overset{iid}{\sim} \mathcal{N}(0, \sigma^2_{i,t}) \tag{2.28}$$

In matrix notation this yields, where I_{T-2} denotes the (T-2)-dimensional identity matrix and T is the sample size,

$$Y^* = \gamma_i \cdot Z^* + \epsilon_i, \quad \epsilon_i \overset{iid}{\sim} \mathcal{N}(0, \sigma^2_{i,t} \cdot I_{T-2}) \tag{2.29}$$

From this (combined with the prior) the posterior distribution follows

$$\gamma_i \,|\, \psi_i, \sigma_{i,t}, \tilde{y}_T, \tilde{z}_T \sim \mathcal{N}(\bar{b}_i, \bar{B}_i)$$

$$\begin{aligned}
\bar{b}_i &= \left(B_i^{-1} + \sigma_i^{-2} Z^{*T} Z^*\right)^{-1} \left(B_i^{-1} b_i + \sigma_i^{-2} Z^{*T} Y^*\right) \\
\bar{B}_i &= \left(B_i^{-1} + \sigma_i^{-2} Z^{*T} Z^*\right)^{-1}
\end{aligned}$$

Then, generating γ_i amounts to computing a draw from this posterior distribution.

2.4.3 Step 3: Generating the economic regime indicator

We consider the model of the business cycle as described in Section 2.1. If we condition this model on the common component, it reduces to an autoregressive model with Markov-switching mean, persistence and variance. We note that, in this set up, the macroeconomic data, $\tilde{y}_T$, carry no information not already captured by the common component, $\tilde{z}_T$. This is the case because the shock processes are independent of each other. The common component, generated in step 1 of the outer algorithm, is therefore treated as data in step 3. We restate the considered model, with the z notation suppressed, for ease of readability.

$$\begin{aligned}
z_t &= \mu_{S_t} + \phi_{S_t} \cdot (z_{t-1} - \mu_{S_{t-1}}) + \sigma_{S_t} \cdot \epsilon_t, && \epsilon_t \overset{iid}{\sim} \mathcal{N}(0,1) \\
\mu_{S_t} &= \mu_0 + \mu_1 \cdot \mathbf{1}_{\{S_t = \text{Recession}\}} && \mu_0 > \mu_0 + \mu_1 \\
\sigma^2_{S_t} &= \sigma^2_0 \cdot (1 + h_\sigma \cdot \mathbf{1}_{\{S_t = \text{Recession}\}}) && h_\sigma > -1
\end{aligned}$$

$$\begin{aligned}
p &= \mathbb{P}[S_t = \text{Recession} \,|\, S_{t-1} = \text{Recession}] \\
q &= \mathbb{P}[S_t = \text{Expansion} \,|\, S_{t-1} = \text{Expansion}]
\end{aligned}$$

Both the parameters of the model and the latent Markov-Switching variable, S_t, are treated as random variables, which we seek to estimate. We therefore need the joint distribution, more specifically the joint posterior distribution, to make inference. Let Θ_z denote the variables associated with the model, except the transition probabilities. Then the joint posterior density is denoted by

$p(\tilde{S}_T, \Theta_z, p, q \mid \tilde{z}_T)$. By the chain rule of probability we can express the joint density as conditional densities instead[17]

$$p(\tilde{S}_T, \Theta_z, p, q \mid \tilde{z}_T) = f(\Theta_z, p, q \mid \tilde{S}_T, \tilde{z}_T) \cdot f(\tilde{S}_T \mid \tilde{z}_T)$$

Now, conditional on $\tilde{S}_T$, the transition probabilities are assumed to be independent of the data and the other parameters of the model

$$= p(\Theta_z \mid \tilde{S}_T, \tilde{z}_T) \cdot p(p, q \mid \tilde{S}_T) \cdot p(\tilde{S}_T \mid \tilde{z}_T)$$

We have now reduced the problem into something we can work with - estimating three conditional distributions. To do so we employ the Gibbs Sampler. We repeat the following three steps until convergence, as done in Kim (1999) [25] Chapter 9,

1. Generate $\tilde{S}_T$ from $p(\tilde{S}_T \mid \tilde{z}_T, \Theta_z, p, q)$
2. Generate p and q from $p(p, q \mid \tilde{S}_T)$
3. Generate Θ_z from $p(\Theta_z \mid \tilde{S}_T, \tilde{z}_T)$

Step 3.1: Generate $\tilde{\mathbf{S}}_\mathbf{T}$ from $\mathbf{p}(\tilde{\mathbf{S}}_\mathbf{T} \mid \tilde{\mathbf{z}}_\mathbf{T}, \mathbf{\Theta}_\mathbf{z}, \mathbf{p}, \mathbf{q})$

We employ the multimove Gibbs Sampler.[18] After the first equality sign, we suppress the conditioning on the model parameters for ease of notation

$$p(\tilde{S}_T \mid \tilde{z}_T, \Theta_z, p, q) = p(S_1, \ldots, S_T \mid \tilde{z}_T)$$

Now, using the chain rule of probability, on the joint distributions

$$\begin{aligned}
&= p(S_T \mid \tilde{z}_T) f(S_1, \ldots, S_{T-1} \mid S_T, \tilde{z}_T) \\
&= p(S_T \mid \tilde{z}_T) f(S_{T-1} \mid S_T, \tilde{z}_T) p(S_1, \ldots, S_{T-2} \mid S_{T-1}, S_T, \tilde{z}_T) \\
&\vdots \\
&= p(S_T \mid \tilde{z}_T) p(S_{T-1} \mid S_T, \tilde{z}_T) p(S_{T-2} \mid S_{T-1}, S_T, \tilde{z}_T) \ldots p(S_1 \mid S_2, \ldots, S_T, \tilde{z}_T)
\end{aligned}$$

Now, using the Markov property of the economic regime indicator, S_t,

$$\begin{aligned}
&= p(S_T \mid \tilde{z}_T) p(S_{T-1} \mid S_T, \tilde{z}_{T-1}) p(S_{T-2} \mid S_{T-1}, \tilde{z}_{T-2}) \ldots p(S_1 \mid S_2, \tilde{z}_1) \\
&= p(S_T \mid \tilde{z}_T) \prod_{t=1}^{T-1} p(S_t \mid S_{t+1}, \tilde{z}_t)
\end{aligned}$$

Thus, the joint distribution of the whole vector, $\tilde{S}_T$, can be obtained by multiplying simpler conditional distributions. We further examine the conditional distribution. Using the chain rule of probability twice, we see

$$p(S_t \mid S_{t+1}, \tilde{z}_t) = \frac{p(S_t, S_{t+1} \mid \tilde{z}_t)}{p(S_{t+1} \mid \tilde{z}_t)} = \frac{p(S_{t+1} \mid S_t, \tilde{z}_t) p(S_t \mid \tilde{z}_t)}{p(S_{t+1} \mid \tilde{z}_t)}$$

[17] The Chain rule of probability: $\mathbb{P}(X, Y) = \mathbb{P}(Y|X) \cdot \mathbb{P}(X)$

[18] Kim (1999) [25] section 9.1.1

Next, using the fact that $\tilde{z}_t$ contain no useful information about S_{t+1} not already captured by S_t,

$$= \frac{p(S_{t+1} \mid S_t)p(S_t \mid \tilde{z}_t)}{p(S_{t+1} \mid \tilde{z}_t)}$$

We disregard the "proportionality constant", as discussed in section 2.3

$$\propto p(S_{t+1} \mid S_t)p(S_t \mid \tilde{z}_t) \tag{2.30}$$

Here $p(S_{t+1} \mid S_t)$ are given by the transition probabilities, p and q, which we have conditioned on, and $p(S_t \mid \tilde{z}_t)$, for $t = 1, \ldots, T$, are generated using Hamilton's basic filter.[19] From the last iteration, $t = T$, S_T is generated. Then for $t = T-1, \ldots, 1$, we generate S_t from equation (2.30). We calculate $\mathbb{P}[S_t = 1 \mid S_{t+1}, \tilde{z}_t]$, generate a random number, u, from a uniform distribution and set

$$S_t = \begin{cases} 1 & \text{if } u \leq \mathbb{P}(S_t = 1 \mid S_{t+1}, \tilde{y}_t) \\ 0 & \text{if } u > \mathbb{P}(S_t = 1 \mid S_{t+1}, \tilde{y}_t) \end{cases} \tag{2.31}$$

Where $S_t = 1$ indicates a recessionary regime and $S_t = 0$ an expansionary regime. We calculate the probability in the following way.

$$\mathbb{P}(S_t = 1 | S_{t+1}, \tilde{z}_t) = \frac{p(S_{t+1} \mid S_t = 1) \cdot p(S_t = 1 \mid \tilde{z}_t)}{\sum_{j=0}^{1} p(S_{t+1} \mid S_t = j) \cdot p(S_t = j \mid \tilde{z}_t)} \tag{2.32}$$

Step 3.2: Generate the transition probabilities

The transition probabilities are assumed to follow beta distributions. A beta distribution is often used to model the distribution of a probability.[20] We therefore use the beta distribution as a conjugate prior and assume the transition probabilities to follow two independent beta distributions, where the hyper parameters, u_{ij}, $i, j = 0, 1$, are known.[21]

$$p \sim \text{Beta}(u_{11}, u_{10})$$
$$q \sim \text{Beta}(u_{00}, u_{01})$$

By the beta density, and as p and q are assumed independent, the joint prior distribution is

$$f(p, q) \propto p^{u_{11}-1}(1-p)^{u_{10}-1}q^{u_{00}-1}(1-q)^{u_{01}-1}$$

We further calculate the Likelihood, where $n_{ij}, i, j = 0, 1$ is obtained by counting the state transitions in, $\tilde{S}_T$, which was generated in the previous step,

$$\text{L}(p, q \mid \tilde{S}_T) = p^{n_{11}}(1-p)^{n_{10}}q^{n_{00}}(1-q)^{n_{01}}$$

Now, the posterior distribution can be computed in closed-form, as we simply add the exponents together to obtain

$$\begin{aligned} f(p, q \mid \tilde{S}_T) &= f(p, q) \cdot \text{L}(p, q \mid \tilde{S}_T) \\ &\propto p^{u_{11}+n_{11}-1}(1-p)^{u_{10}+n_{10}-1}q^{u_{00}+n_{00}-1}(1-q)^{u_{01}+n_{01}-1} \end{aligned}$$

[19] See Hamilton (1989) [15] for further explanation on Hamilton's Basic Filter

[20] As the domain of the beta distribution can be viewed as a probability - see [44]

[21] A conjugate prior is a prior probability chosen such that the posterior ends up being in the same distribution as the prior.

Thus, obtaining the two independent posterior distributions

$$p \,|\, \tilde{S}_T \sim \text{Beta}(u_{11} + n_{11}, u_{10} + n_{10})$$
$$q \,|\, \tilde{S}_T \sim \text{Beta}(u_{00} + n_{00}, u_{01} + n_{01})$$

Then, to generate the transition probabilities, p and q, we draw from these posterior distributions.

Step 3.3: Generate the common component parameters

In this step we seek to generate the Markov-switching mean, persistence and variance associated with the common component. The idea is then

1. Generate ϕ conditional on the other parameters, $\tilde{S}_T$ and $\tilde{z}_T$.
2. Generate μ_0 and μ_1 conditional on the other parameters, $\tilde{S}_T$ and $\tilde{z}_T$.
3. Generate σ_0^2 conditional on the other parameters, $\tilde{S}_T$ and $\tilde{z}_T$.
4. Generate h_σ conditional on the other parameters, $\tilde{S}_T$ and $\tilde{z}_T$.

Step 3.3.1: Generate ϕ

We consider the model equation, restated for ease of readability,

$$z_t = \mu_{S_t} + \phi \cdot (z_{t-1} - \mu_{S_{t-1}}) + \sigma_{S_t} \cdot \epsilon_t, \qquad \epsilon_t \overset{iid}{\sim} \mathcal{N}(0,1)$$

We rearrange and divide both sides of the equation with σ_{S_t} to obtain

$$\frac{z_t - \mu_{S_t}}{\sigma_{S_t}} = \phi \cdot \frac{z_{t-1} - \mu_{S_{t-1}}}{\sigma_{S_t}} + \epsilon_t, \qquad \epsilon_t \overset{iid}{\sim} \mathcal{N}(0,1)$$

In matrix notation this yields, where $Z^*_{shifted}$ defines the Z^* matrix shifted by one time period and I_T is the T-dimensional identity matrix,

$$Z^* = \phi \cdot Z^*_{shifted} + \epsilon_t, \qquad \epsilon_t \sim \mathcal{N}(0, I_T)$$

We assume that ϕ has a normal prior with known initial hyper parameters

$$\phi \,|\, \sigma_0, \mu_0, \mu_1, h_\sigma \sim \mathcal{N}(b_0, B_0)$$

The posterior distribution is then.

$$\phi \,|\, \sigma_0, \mu_0, \mu_1, h_\sigma, \tilde{S}_T, \tilde{z}_T \sim \mathcal{N}(b_2, B_2)$$

$$b_2 = \left(B_0^{-1} + {Z^*_{shifted}}^T Z^*_{shifted}\right)^{-1} \left(B_0^{-1} b_0 + {Z^*_{shifted}}^T Z^*\right)$$
$$B_2 = \left(B_0^{-1} + {Z^*_{shifted}}^T Z^*_{shifted}\right)^{-1}$$

Then to generate ϕ we draw one sample from this posterior distribution.

Step 3.3.2: Generate μ_0, μ_1

We follow the steps of Kim (1999) [25] section 9.1.3 and merge the model equations to one

$$z_t - \phi_{S_t} \cdot z_{t-1} = \mu_0 \cdot (1 - \phi_{S_t}) + \mu_1 \cdot (\mathbf{1}_{\{S_t=\text{Recession}\}} - \phi_{S_t} \cdot \mathbf{1}_{\{S_{t-1}=\text{Recession}\}}) + \epsilon_t^\sigma, \qquad \epsilon_t^\sigma \overset{iid}{\sim} \mathcal{N}(0, \sigma_{S_t}^2)$$

Divide both sides of the equation with σ_{S_t}

$$\frac{z_t - \phi_{S_t} \cdot z_{t-1}}{\sigma_{S_t}} = \frac{\mu_0 \cdot (1 - \phi_{S_t}) + \mu_1 \cdot (\mathbf{1}_{\{S_t=\text{Recession}\}} - \phi_{S_t} \cdot \mathbf{1}_{\{S_{t-1}=\text{Recession}\}})}{\sigma_{S_t}} + \epsilon_t, \qquad \epsilon_t \overset{iid}{\sim} \mathcal{N}(0, 1)$$

Define

$$\begin{aligned} z_t^* &= \frac{z_t - \phi_{S_t} \cdot z_{t-1}}{\sigma_{S_t}} \\ x_0^* &= \frac{1 - \phi_{S_t}}{\sigma_{S_t}} \\ x_1^* &= \frac{\mathbf{1}_{\{S_t=\text{Recession}\}} - \phi_{S_t} \cdot \mathbf{1}_{\{S_{t-1}=\text{Recession}\}}}{\sigma_{S_t}} \end{aligned}$$

Then

$$z_t^* = \mu_0 \cdot x_0^* + \mu_1 \cdot x_1^* + \epsilon_t, \qquad \epsilon_t \overset{iid}{\sim} \mathcal{N}(0, 1)$$

And in matrix notation, where $\tilde{\mu} = [\mu_0, \mu_1]^T$ and I_T is the T-dimensional identity matrix,

$$Z^* = X^* \tilde{\mu} + \varepsilon, \qquad \varepsilon \sim \mathcal{N}(0, I_T)$$

As we have the model framework in place, we are ready for Bayesian inference. We assume the parameters to have a normal prior distribution, where the initial hyper parameters are known

$$\tilde{\mu} \,|\, \sigma_0^2, \sigma_1^2 \sim \mathcal{N}(b_0, B_0)$$

The posterior distribution is then.

$$\begin{aligned} \tilde{\mu} \,|\, \sigma_0^2, \sigma_1^2, \tilde{S}_T, \tilde{z}_T &\sim \mathcal{N}(b_1, B_1) \\ b_1 &= \left(B_0^{-1} + X^{*T} X^*\right)^{-1} \left(B_0^{-1} b_0 + X^{*T} Z^*\right) \\ B_1 &= \left(B_0^{-1} + X^{*T} X^*\right)^{-1} \end{aligned}$$

Then, to generate $\tilde{\mu}$, we draw from this posterior distribution, and if they meet the model criteria, $\mu_0 > \mu_0 + \mu_1$, we keep the sample. Otherwise we draw again.

Step 3.3.3: Generate σ_0^2

We remember the variance equation (2.5), $\sigma_{S_t}^2 = \sigma_0^2 \cdot (1 + h_\sigma \cdot \mathbf{1}_{\{S_t=\text{Recession}\}})$. We can see that the variance in the recessionary regime is given by $\sigma_1^2 = \sigma_0^2(1 + h_\sigma)$. Following the same idea as in the previous step, we merge the model equations to one, but this time we divide with $\sqrt{1 + h_\sigma \cdot \mathbf{1}_{\{S_t=\text{Recession}\}}}$ instead, to obtain

$$z_t^* = \mu_0 \cdot x_0^* + \mu_1 \cdot x_1^* + \epsilon_t^0, \qquad \epsilon_t^0 \overset{iid}{\sim} \mathcal{N}(0, \sigma_0^2)$$

Where

$$
\begin{aligned}
z_t^* &= \frac{z_t - \phi_{S_t} \cdot z_{t-1}}{\sqrt{1 + h_\sigma \cdot \mathbf{1}_{\{S_t = \text{Recession}\}}}} \\
x_0^* &= \frac{1 - \phi_{S_t}}{\sqrt{1 + h_\sigma \cdot \mathbf{1}_{\{S_t = \text{Recession}\}}}} \\
x_1^* &= \frac{\mathbf{1}_{\{S_t = \text{Recession}\}} - \phi_{S_t} \cdot \mathbf{1}_{\{S_{t-1} = \text{Recession}\}}}{\sqrt{1 + h_\sigma \cdot \mathbf{1}_{\{S_t = \text{Recession}\}}}} \\
\epsilon_t^0 &= \frac{\epsilon_t^\sigma}{\sqrt{1 + h_\sigma \cdot \mathbf{1}_{\{S_t = \text{Recession}\}}}}
\end{aligned}
$$

As we seek to infer on variance, we choose an inverse gamma distribution, with known initial hyper parameters, as a conjugate prior, as done in previous sections.

$$
\sigma_0^2 \,|\, h_\sigma, \mu_0, \mu_1, \phi \sim IG\left(\frac{c_0}{2}, \frac{C_0}{2}\right)
$$

From this the posterior distribution follows, where T is the sample size.

$$
\sigma_0^2 \,|\, h_\sigma, \mu_0, \mu_1, \phi, \tilde{S}_T, \tilde{z}_T \sim IG\left(\frac{c_1}{2}, \frac{C_1}{2}\right)
$$

$$
\begin{aligned}
c_1 &= c_0 + T \\
C_1 &= C_0 + \sum_{t=1}^{T} \left(z_t^* - \mu_0 \cdot x_0^* - \mu_1 \cdot x_1^*\right)^2
\end{aligned}
$$

Then to generate σ_0^2 we draw from this posterior distribution.

Step 3.3.4: Generate $\mathbf{h}_\sigma$

We employ the same methodology as in the previous two sections and merge the business cycle model into one equation. However this time we divide by σ_0 to obtain

$$
z_t^* = \mu_0 \cdot x_0^* + \mu_1 \cdot x_1^* + \epsilon_t^1, \qquad \epsilon_t^1 \overset{iid}{\sim} \mathcal{N}(0, 1 + h_\sigma \cdot \mathbf{1}_{\{S_t = \text{Recession}\}})
$$

Where

$$
\begin{aligned}
z_t^* &= \frac{z_t - \phi_{S_t} \cdot z_{t-1}}{\sigma_0} \\
x_0^* &= \frac{1 - \phi_{S_t}}{\sigma_0} \\
x_1^* &= \frac{\mathbf{1}_{\{S_t = \text{Recession}\}} - \phi_{S_t} \cdot \mathbf{1}_{\{S_{t-1} = \text{Recession}\}}}{\sigma_0} \\
\epsilon_t^0 &= \frac{\epsilon_t^\sigma}{\sigma_0}
\end{aligned}
$$

Define $\hat{h} = 1 + h_\sigma$. We then assume an inverse gamma conjugate prior, with known initial hyper parameters, for $\hat{h}$.

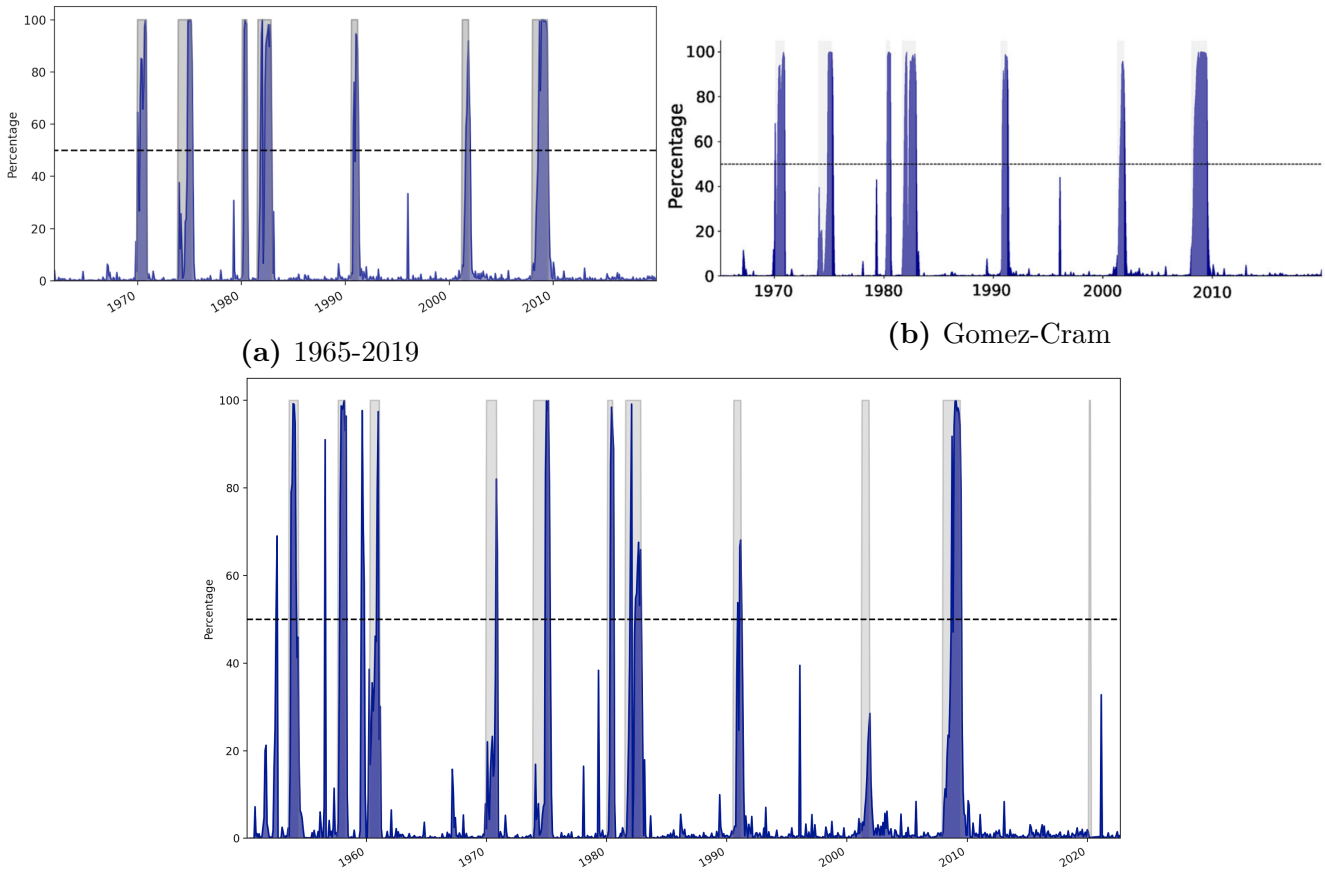
$$\hat{h} \,|\, \sigma_0^2, \mu_0, \mu_1, \phi \sim IG\left(\frac{c_0}{2}, \frac{C_0}{2}\right)$$

We note that h_σ only depend on the recessionary regime. We therefore create a temporary index, where we only consider the cases where $S_t = \text{Recession}$. Let $R_1 = \{t : S_t = \text{Recession}\}$ and let T_1 be the sample size of this index. Then the posterior distribution follows

$$\hat{h} \,|\, \sigma_0^2, \mu_0, \mu_1, \phi, \tilde{S}_T, \tilde{z}_T \sim IG\left(\frac{c_2}{2}, \frac{C_2}{2}\right)$$

$$\begin{aligned} c_2 &= c_0 + T_1 \\ C_2 &= C_0 + \sum_{R_1} \left(z_t^* - \mu_0 \cdot x_0^* - \mu_1 \cdot x_1^*\right)^2 \end{aligned}$$

Then to generate h_σ, we draw from this posterior distribution, and if the draw follows the model criteria, $h_\sigma > -1$, we keep the sample or otherwise draw again.



(a) 1965-2019

(b) Gomez-Cram

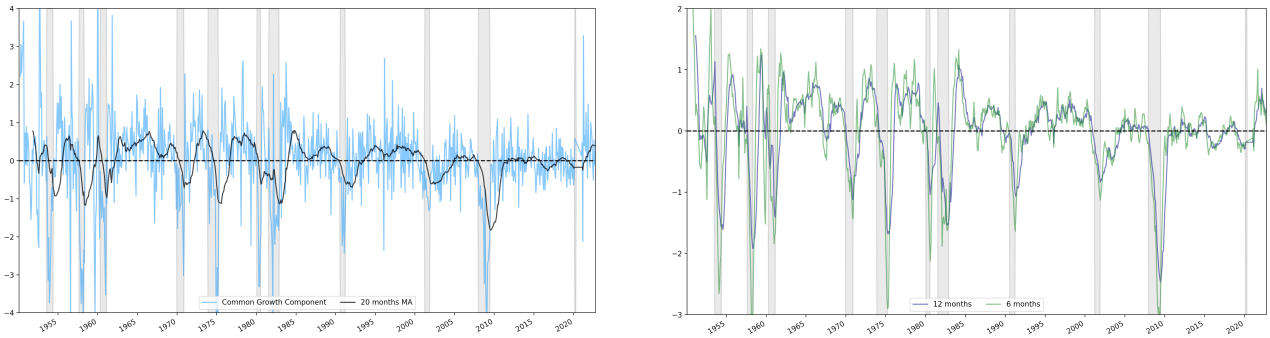
(c) 1950-2022

Figure 2.2: Filtered probability of recession, $\hat{\pi}_{t|t}$. *The plot shows the posterior median estimate (navy). The grey bars indicate NBER recessions. The lower panel shows the recession probability for the full sample, January 1950 to August 2022, with dates from March 2020 to November 2020 removed. The right panel shows a screenshot of Figure 1 from Gomez-Cram (2022) [14] for comparison. The left panel shows the plot generated from this thesis using the same dates, January 1965 to December 2019.*

Posterior Inference Results

From Figure 2.2 we find, firstly, that the recession probabilities strongly coincide with those of Gomez-Cram (2022) [14]. Secondly, in Figure 2.2a, we find that the filtered probability of recession significantly coincides with the NBER recession dates. In non-recessionary times the probabilities are close to zero. At the onset of each recession the probability rises sharply, and almost exactly at the NBER end date quickly drops to almost zero again. Interestingly, from Figure 2.2, we see that, in both sample periods, our model best captures the earlier recessions, as well as the 1982 recession. Compared to these, the recession probabilities in the following recessions are less pronounced. Especially the 1900 and 2001 recessions have had predictability issues in previous literature. E.g. Kauppi and Saikkonen (2008) [18], using probit models with lagged dependent variables, shows that the 1990 and 2001 recessions were difficult to predict. Comparing Figure 2.2a and Figure 2.2c, we find that when expanding the sample period to begin 1950 instead of 1965, the model's ability to capture the later recession decreases - especially the recession probabilities of the 2001 recession. After expanding the sample period the values drop below the 50% threshold as

indicated by the black dotted line. This indicates that the model might benefit from computing the recession probabilities in sub-sample periods and thereafter collecting them into one data frame.



(a) Common Component **(b)** Moving averages

Figure 2.3: Common component. *The sample period runs from January 1950 to August 2022, with dates 2020:3-2020:11 removed. The grey bars indicate NBER recessions.*

Figure 2.3a shows the posterior median estimate of the common component, as introduced in equation 2.3 and derived in section 2.4.1. The common component is interpreted as our real-time belief about the business cycle. We see that this is volatile, however clearly varies with the business cycle. This fact is clearly seen in the moving averages in Figure 2.3b. Considering the 6- and 12-months moving averages (green and blue respectively), we see how these drop significantly below zero almost exactly at the first NBER recession date (as indicated by the grey shaded areas) and return above/around zero again after the NBER recession end date.

Chapter 3

Expected excess return

In this chapter we employ the business cycle estimates to predict the equity premium. We obtain two one month ahead forecasts of excess return, which employs the same filtered probabilities of recession, however exploit them differently - one indirectly through a state-space model and the other one directly in a Ordinary Least Square regression framework. We begin with the more sophisticated state-space-based forecast as introduced by Gomez-Cram (2022) [14].

3.1 Model set up

We consider the following model for excess stock return. We assume a regime-switching model, where conditional expected return is treated as a latent variable. Let r^e_{t+1} define the excess market stock return. This variable is observable and depend on the latent variable. Let μ_{t+1} denote the conditional expected excess return. We assume this to follow a first order autoregressive process. The described model takes the following form

$$r^e_{t+1} = \mu_t + \epsilon^r_{t+1} \tag{3.1}$$

$$\mu_{t+1} = \mu_0 + \rho \cdot (\mu_t - \mu_0) + \epsilon^\mu_{t+1} \tag{3.2}$$

Here μ_0 and ρ denotes the unconditional mean and persistence of μ_{t+1}. Both variables further depend on shock processes, ϵ^r_{t+1} and ϵ^μ_{t+1}, where the first captures unexpected excess return, and the latter represents shocks to expected returns. The two shocks are assumed to be conditionally Gaussian and depend on the state of the economy. We assume that the correlation between shocks, $\rho_{r,\mu}$, is stable across the different states.[1] S_t is the economic regime indicator, which tells us whether the economy is in an expansionary or recessionary period. This is discussed and generated in the previous chapter on the business cycle.

$$(\epsilon^r_{t+1}, \epsilon^\mu_{t+1})^T \overset{iid}{\sim} \mathcal{N}(0, \Sigma_{S_t}), \qquad \Sigma_{S_t} = \begin{bmatrix} \sigma^2_{r,S_t} & \sigma^2_{r\mu,S_t} \\ \sigma^2_{\mu r,S_t} & \sigma^2_{\mu,S_t} \end{bmatrix}, \qquad \sigma^2_{r\mu,S_t} = \sigma^2_{\mu r,S_t} = \rho_{r,\mu} \cdot \sigma_{r,S_t} \cdot \sigma_{\mu,S_t}$$

We further define the fraction of variance in excess returns explained by the conditional expected excess return.

$$\phi_{S_t} = \frac{\mathbb{V}[\mu_t \mid S_t]}{\mathbb{V}[r^e_t \mid S_t]} \tag{3.3}$$

[1] Gomez-Cram (2022) [14] tests different model alternatives in his Appendix and find that this specification is preferred over the model alternatives.

3.1.1 Expected excess returns estimate

One of the main outputs of this thesis is an estimate of the one-month-ahead excess return forecast, $\hat{\mu}_{t+1|t}$. Let Θ_j denote the vector of model parameters associated with the economic regime j, $j \in \{\text{Expansion}, \text{Recession}\}$, and Θ be the collection of $\Theta_j's$. Further let $\tilde{r}_t^e$ denote the observed excess returns up to time t, $\tilde{r}_t^e = [r_1^e, \ldots, r_t^e]$. We wish to obtain the predictive density of excess return, $p(r_{t+1}^e \,|\, \Theta_j, \tilde{r}_t^e)$, as the predictive density is used for making inference about a new data point.[2] This is obtained by conditioning model (3.1) and (3.2) on the model parameters and excess return data. Then, the aggregate predictive density of returns is given by a weighted average of the two predictive densities. The weights are given by the probability of being in a recession at time t conditional on the information available at time t, $\hat{\pi}_{t|t} = \mathbb{E}[S_t = \text{Recession} \,|\mathcal{F}_t]$, and thus time-varying. This quantity was introduced in section 2.2. This idea takes inspiration in the extensive literature on density forecast combinations. For example the following two articles on prediction pools from Journal of Econometrics: 1) Geweke and Amisano (2011) [13], who combine the predictive densities of alternative models using optimal linear weights; and 2) Schorfheide et. al. (2016) [27] who combine the predictive densities from two DSGE models. We obtain the predictive density of excess return.

$$p(r_{t+1}^e \,|\, \Theta, \tilde{r}_t^e) = (1 - \hat{\pi}_{t|t}) \cdot p(r_{t+1}^e \,|\, \Theta_{\text{Expansion}}, \tilde{r}_t^e) + \hat{\pi}_{t|t} \cdot p(r_{t+1}^e \,|\, \Theta_{\text{Recession}}, \tilde{r}_t^e) \tag{3.4}$$

Using this we define the filtered expected returns, given some model parameters.

$$\hat{\mu}_{t+1|t}(\Theta) := \mathbb{E}[r_{t+1}^e \,|\, \Theta, \tilde{r}_t^e] = \int r_{t+1}^e p(r_{t+1}^e \,|\, \Theta, \tilde{r}_t^e)\, dr_{t+1}^e \tag{3.5}$$

We can discard the Θ argument by integrating over the posterior distribution of Θ, $p(\Theta \,|\, \tilde{r}_t^e)$, and in this way eliminate dependence on the model parameters in our prediction of the new data point. The posterior density of a given set of parameters, $\Theta(\omega)$, tells us about the credibility of these exact parameters. In this way, the larger credibility a specific set of parameters has, the higher weight $\hat{\mu}_{t+1|t}(\Theta(\omega))$ will have in the final estimate. We thereby obtain an estimate of expected returns, which was what we wanted.

$$\begin{aligned} \hat{\mu}_{t+1|t} := \mathbb{E}[r_{t+1}^e \,|\, \tilde{r}_t^e] &= \int \hat{\mu}_{t+1|t}(\Theta) p(\Theta \,|\, \tilde{r}_t^e)\, d\Theta \\ &= \int \int r_{t+1}^e p(r_{t+1}^e \,|\, \Theta, \tilde{r}_t^e) p(\Theta \,|\, \tilde{r}_t^e)\, d\Theta\, dr_{t+1}^e \end{aligned} \tag{3.6}$$

Often it will not be possible to compute this in closed form, as high dimensional integrals quickly becomes too complex to deal with. In such cases we will instead embrace simulation-based inference, such as Markov Chain Monte Carlo methods, as done in this thesis. To do so we need the state-space representation of the model.

3.1.2 State-space representation

The state-space representation is in the following form.

$$\text{(Measurement equation)} \quad y_{t+1} = A_{t+1} \cdot (H_0 + H_1 \cdot \beta_{t+1} + \nu_{t+1}), \qquad \nu_{t+1} \overset{iid}{\sim} \mathcal{N}(0, R) \tag{3.7}$$

$$\text{(State-transitions equation)} \quad \beta_{t+1} = F_{0,S_t,S_{t+1}} + F_{1,S_{t+1}} \cdot \beta_t + \omega_{t+1}, \qquad \omega_{t+1} \overset{iid}{\sim} \mathcal{N}(0, Q_{S_{t+1}}) \tag{3.8}$$

[2] See e.g. M. Sugiyama (2016) [41] Chapter 17 for a description of predictive density. Note that this differs from the posterior density.

The model (3.1) and (3.2) can be expressed in state-space form using the following matrices.

$$A_{t+1} = 1 \quad H_0 = 0 \quad H_1 = \begin{bmatrix} 0 & 1 & 1 \end{bmatrix} \quad R = 0$$

Given β_{t+1} the matrices associated with the state-transition equation are

$$\beta_{t+1} = \begin{bmatrix} \mu_{t+1} \\ \mu_t \\ \epsilon^r_{t+1} \end{bmatrix}, \quad F_0 = \begin{bmatrix} \mu_0(1-\rho) \\ 0 \\ 0 \end{bmatrix} \quad F_1 = \begin{bmatrix} \rho & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad Q_{S_{t+1}} = \begin{bmatrix} \sigma^2_{r,S_t} & 0 & \sigma^2_{r\mu,S_t} \\ 0 & 0 & 0 \\ \sigma^2_{r\mu,S_t} & 0 & \sigma^2_{\mu,S_t} \end{bmatrix}$$

3.2 Posterior Inference

As the economic state indicator, S_t, has been generated in the previous step, more specifically section 2.4.3, we can condition on it, and employ the Gibbs sampler in order to obtain posterior inference of the expected returns.[3] Let N define the number of simulations in the Gibbs sampler.

The algorithm

0. Initialize $\Theta^{(0)}$. We follow Gomez-Cram (2022) [14] and initialize at the posterior mode.

   For $i = 1, \ldots, N$:

1. Draw $\tilde{\mu}_T^{(i)}$ conditional on the parameters, $\Theta^{(i-1)}$, and the return data, $\tilde{y}_T$. We employ the Kalman filter of Section 3.2.1.

2. Draw parameters, $\Theta^{(i)}$, conditional on $\tilde{\mu}_T^{(i)}$ and the data $\tilde{y}_T$. We employ the Metropolis-Hastings algorithm of Section 3.2.2.

The initial values and priors used are the same as Gomez-Cram (2022) [14] and can be seen in Table 3.2, together with the posterior inference results. We choose vague priors to decrease the influence on the posterior estimates, for example ϕ_{S_t}, the fraction of variance explained by the expected excess return, as specified in equation 3.3. We would expect this quantity to vary across regimes, however we define the priors on the same interval, $[0, 0.2]$, thus allowing expected returns to be equally volatile in the recessionary and expansionary regime.

3.2.1 The Kalman filter

We observe that, conditional on S_t, the state-space model (3.7) and (3.8) is a linear Gaussian model. This can be seen as the observations depend linearly on the state variable, the state variable depends linearly on the value of the previous state variable, and the innovations are Gaussian. Thus, by Lemma 1, the conditional distributions are also Gaussian, and we can therefore apply the Gaussian Kalman filter. In the previous chapter we used the Kalman filter to infer on the non-observable common component. In this section we wish to employ it to infer on the non-observable conditional expected excess return. We recall that the algorithm works in two steps - prediction and updating - where both steps are performed at time $t+1$ and which step depend on the information set.

[3] The Gibbs sampler is described in section 2.3.1.

Prediction: We have not yet observed the market excess return data, y_{t+1}, but wish to infer on this given the available information. We call this predictor, $y_{t+1|t}$. $P_{t+1|t}$ is the variance associated with the non-observable variable we wish to estimate.

$$\beta_{t+1|t} = F_0 + F_1 \cdot \beta_{t|t} \tag{3.9}$$
$$P_{t+1|t} = F_1 \cdot \mu_{t|t} \cdot F_1^T + Q \tag{3.10}$$
$$y_{t+1|t} = A \cdot (H_0 + H_1 \cdot \beta_{t+1|t}) \tag{3.11}$$

The new observation, y_{t+1}, is made. The prediction error and the variance relating is calculated.

$$\eta_{t+1|t} = y_{t+1} - y_{t+1|t} \tag{3.12}$$
$$f_{t+1|t} = (A \cdot H_1) \cdot P_{t+1|t} \cdot (A \cdot H_1)^T + R \tag{3.13}$$

Updating: Incorporating the new data point yields, where the weight given to the new information is determined by the Kalman gain,

$$\beta_{t+1|t+1} = \beta_{t+1|t} + \underbrace{P_{t+1|t} \cdot (A \cdot H_1)^T \cdot f_{t+1|t}^{-1}}_{\text{Kalman gain}} \cdot \eta_{t+1|t} \tag{3.14}$$
$$P_{t+1|t+1} = P_{t+1|t} + P_{t+1|t} \cdot (A \cdot H_1)^T \cdot f_{t+1|t}^{-1} \cdot (A \cdot H_1) \cdot P_{t+1|t} \tag{3.15}$$

Intuitive explanation

We can write out the Kalman equations by employing the matrices in the state-space representation. The measurement equation collapses to $y_{t+1} = r_{t+1}^e$. Now, imagine we are standing in the beginning of month $t+1$. We currently have the information up to and including month t. We have observed the regime indicator, i.e. conditional on regime j, $j \in \{\text{Expansion}, \text{Recession}\}$, and wish to infer what the expected excess return might be. The expected excess return, μ_{t+1}, conditional the information available at time t and on regime j, is[4]

Prediction:

$$\begin{aligned}
\mu_{t+1|t,j} &= \mathbb{E}[\mu_{t+1} \,|\, \Theta_j, \tilde{r}_t^e, S_t = j] \\
&= F_0 + F_1 \cdot \mu_{t|t,j} \\
&= \mu_0 \cdot (1-\rho) + \rho \cdot \mu_{t|t,j} \\
&= \mu_0 + \rho \cdot (\mu_{t|t,j} - \mu_0)
\end{aligned}$$

Now imagine we are in the mid/end of month $t+1$. The new excess return information, r_{t+1}^e, has been released and we wish to update our belief on the conditional expected excess return, $\mu_{t+1|t,j}$. The prediction error is used to update our belief. The weight which we assign to this new information is determined by the Kalman gain. This quantity depends on the uncertainty of the old belief and the uncertainty of the new information.

Updating:

$$\begin{aligned}
\mu_{t+1|t+1,j} &= \mathbb{E}[\mu_{t+1} \,|\, \Theta_j, \tilde{r}_{t+1}^e, S_t = j] \\
&= \mu_{t+1|t} + \underbrace{P_{t+1|t} \cdot (A \cdot H_1)^T \cdot f_{t+1|t}^{-1}}_{\text{Kalman gain}} \cdot (r_{t+1}^e - \mu_{t+1|t,j})
\end{aligned}$$

[4]Note that this is one of the components in what we refer to as the 1-month ahead excess return forecast.

The Kalman gain

We can further build on the intuitive understanding of the Kalman gain, on the basis of the discussion in section 2.4.1. Here we deduced how the Kalman gain depends positively on the uncertainty surrounding the expected excess return given the information up to time t, $P_{t+1|t,j}$, and negatively on the uncertainty surrounding shocks to excess return, R. This is seen in Table 3.1, which reports the posterior distribution of the Kalman Gain during the expansionary and recessionary regime.[5] We find that excess return news are negatively weighted in the expansionary regime and positively weighted in the recessionary. Intuitively this makes sense. Assume we experience a lower than expected excess return (meaning the prediction error, $\eta_{t+1|t}$, is negative). In the recessionary regime this is more likely to be due to a shock to excess return (large R). The Kalman gain is negative, thereby causing a negative prediction error to lead to an increase in the forecast of future returns. In the expansionary period, if expected returns are persistent, as Gomez-Cram (2022) [14] argue, then the Kalman gain will be positive and thus update future returns in the same direction as past returns.

Table 3.1: *Kalman gain in the recessionary and expansionary regime. The sample period runs from January 1950 to December 2019.*

	Kalman Gain		
	5%	50%	95%
Expansion	-0.01405	-0.01320	-0.00852
Recession	0.05868	0.06156	0.06432

3.2.2 Metropolis-Hasting algorithm

To estimate the model parameters we employ the Metropolis-Hastings (M-H) algorithm. This enables us to compute samples from any distribution, if we can compute a proportional value, thus removing the need to compute the often computationally heavy normalization factor. The Metropolis-Hastings algorithm is a Markov Chain Monte Carlo (MCMC) method. MCMC was discussed in section 2.3.1. The idea behind MCMC is to design a specific type of Markov chain which converges to the target distribution. The design part lies in the choice of a kernel density to simulate the Markov chain. The Metropolis-Hastings algorithm is a specific design of this, as was the Gibbs sampling approach we applied in the previous chapter.

Suppose we wish to compute a sample from the distribution, $p(x)$, which is equal to some known distribution, $f(x)$, divided by an unknown normalization constant, C, i.e. $p(x) = \frac{f(x)}{C}$. The goal

[5] These quantities are computed when running the Kalman Filter for posterior inference.

is then to obtain samples from the unknown distribution, $p(x)$, using only the known distribution, $f(x)$. The M-H algorithm works in two steps

1. Propose a candidate for the next variable in the Markov Chain, by sampling from an easy distribution conditional on the current state variable, fx. the Gaussian, $g(x_{t+1}|x_t) = \mathcal{N}(x_t, \Sigma)$.

2. Accept the candidate with probability $A(x_t, x_{t+1})$. We call this the acceptance probability.

The design must satisfy the detailed balance condition, that is the probability of being in state x_t and moving to state x_{t+1} must be the same as the probability of being in state x_{t+1} and moving to state x_t. Let $T(x_t, x_{t+1})$ be a kernel that defines a Markov chain whose stationary distribution is $p(\cdot)$. Then, $\forall t$, the detailed balance condition states

$$p(x_t) \cdot T(x_t, x_{t+1}) = p(x_{t+1}) \cdot T(x_{t+1}, x_t)$$

The density kernel $T(x_t, x_{t+1})$ consists of a transition kernel $g(x_{t+1} \,|\, x_t)$ and of a probability $A(x_t, x_{t+1})$. In more intuitive words, the probability of transitioning from state x_t to x_{t+1}, consists of the probability of proposing the candidate (step 1) and the probability of accepting the candidate (step 2).

$$p(x_t) \cdot g(x_{t+1} \,|\, x_t) \cdot A(x_t, x_{t+1}) = p(x_{t+1}) \cdot g(x_t \,|\, x_{t+1}) \cdot A(x_{t+1}, x_t)$$

Rearranging this we find

$$\frac{A(x_t, x_{t+1})}{A(x_{t+1}, x_t)} = \frac{p(x_{t+1})}{p(x_t)} \cdot \frac{g(x_t \,|\, x_{t+1})}{g(x_{t+1} \,|\, x_t)}$$

Metropolis-Hastings defines the probability of accepting a new proposal in the following way

$$A^{MH}(x_t, x_{t+1}) = \min\left(1, \frac{p(x_{t+1})}{p(x_t)} \cdot \frac{g(x_t \,|\, x_{t+1})}{g(x_{t+1} \,|\, x_t)}\right)$$

One exceptionally useful feature of this, is that we do not need to be able to compute the normalization constant, as it cancels out in the fraction. This is an important feature when conducting Bayesian Inference, as computing the denominator of the posterior distribution (the marginal likelihood) often is difficult or computationally impossible. Now, what is left, is to prove that this M-H chain is reversible with stationary distribution $p(\cdot)$. We do this by showing the detailed balance condition holds.[6]We start by examining the case where the probability of acceptance is less that 1, i.e.

$$A^{MH}(x_t, x_{t+1}) = \frac{p(x_{t+1})}{p(x_t)} \cdot \frac{g(x_t \,|\, x_{t+1})}{g(x_{t+1} \,|\, x_t)} \quad \text{and} \quad A^{MH}(x_{t+1}, x_t) = 1$$

We expand on the left hand side of the detailed balance condition,

$$\begin{aligned}
p(x_t) \cdot T(x_t, x_{t+1}) &= p(x_t) \cdot g(x_{t+1} \,|\, x_t) \cdot A^{MH}(x_t, x_{t+1}) \\
&= p(x_t) \cdot g(x_{t+1} \,|\, x_t) \cdot \frac{p(x_{t+1})}{p(x_t)} \cdot \frac{g(x_t \,|\, x_{t+1})}{g(x_{t+1} \,|\, x_t)} \\
&= p(x_{t+1}) \cdot g(x_t \,|\, x_{t+1}) \cdot 1 \\
&= p(x_{t+1}) \cdot g(x_t \,|\, x_{t+1}) \cdot A^{MH}(x_{t+1}, x_t) \\
&= p(x_{t+1}) \cdot T(x_{t+1}, x_t)
\end{aligned}$$

Which was what we wanted. We have showed that it holds for $p(x_{t+1})g(x_t \mid x_{t+1}) \leq p(x_t)g(x_{t+1} \mid x_t)$. By symmetry it further holds for $p(x_{t+1})g(x_t \mid x_{t+1}) > p(x_t)g(x_{t+1} \mid x_t)$, thus we can conclude that $p(\cdot)$ is the stationary distribution of the Markov chain with density kernel $T(x_t, x_{t+1})$. [7]

In order to apply the algorithm, the next choice is the design of the transition kernel, $g(x_{t+1} \mid x_t)$. In this thesis we will choose the random walk, as originally proposed by Metropolis (1953) [32], meaning that a new candidate, x^*_{t+1}, is generated under the assumption

$$x^*_{t+1} = x_{t+1} = x_t + \omega_{t+1}, \quad \omega_{t+1} \sim \mathcal{N}(0, \Omega) \quad \Rightarrow \quad x^*_{t+1} \sim \mathcal{N}(x_t, \Omega)$$

For the choice of Ω we follow Gomez-Cram (2022) [14]. The Metropolis-Hastings Random Walk algorithm (MH-RW) is then,[8]

1. Set initial values, x_0.
2. For $t = 1, \ldots, S$, where S is *sufficiently* large,
   a. Generate a proposed state, x^*_{t+1}, from the transition kernel,
   $$g(x_{t+1} \mid x_t) = N(x_t, \Omega)$$
   b. Compute the probability of accepting the proposed candidate,
   $$A^{MH}(x_t, x^*_{t+1}) = \min\left(1, \frac{p(x^*_{t+1})}{p(x_t)} \cdot \frac{g(x_t \mid x^*_{t+1})}{g(x^*_{t+1} \mid x_t)}\right)$$
   c. Draw a single value, u, from $U(0, 1)$. If the candidate is rejected, we save the previous.
   $$x_{t+1} = \begin{cases} x^*_{t+1} & A^{MH}(x_t, x^*_{t+1}) > u \\ x_t & A^{MH}(x_t, x^*_{t+1}) \leq u \end{cases}$$

The intuition behind the algorithm is pretty cool. Assume for a moment that we're in the symmetric case, then

$$A^{MH}(x_t, x^*_{t+1}) = \min\left(1, \frac{p(x_{t+1})}{p(x_t)}\right)$$

There are two cases: 1) The proposed candidate is a more probable state (according to the target distribution). Then the transition probability will equal 1 and the new candidate will for certain be accepted. This allows us to move incrementally closer to more probability dense areas, and thus providing more samples here. 2) The proposed candidate is a less probable state (according to the target distribution). In this case we still might accept the candidate, however the bigger the distance to the candidate, the less probability of acceptance. In this way, we still obtain samples in the tails, but not too many. This means that even if we are at the maximum/most probable

[6] Following the steps of Martin et al. (2022) [33]

[7] If $p(x_{t+1})g(x_t \mid x_{t+1}) > p(x_t)g(x_{t+1} \mid x_t)$ then $A^{MH}(x_t, x_{t+1}) = 1$ and $A^{MH}(x_{t+1}, x_t) = \frac{1}{\frac{p(x_{t+1})}{p(x_t)} \cdot \frac{g(x_t \mid x_{t+1})}{g(x_{t+1} \mid x_t)}}$, so by symmetry the same arguments works.

[7] For further explanation we refer to the referenced Journal of Finance article [14].

[8] K. Triantafyllopoulos (2021) [43]

parameter value, the algorithm still has positive probability of moving to a less probability dense area. Hence, this is not an optimization algorithm. Our goal is to obtain inference, i.e. to explore the whole target distribution - it is not to find the single parameter with the highest possible probability.

Further, it is important to note that MH-RW is a first order MCMC method, meaning that the samples generated are dependent upon each other, due to the incremental behaviour of the Markov chain. This means that samples in the beginning of the chain may depend heavily on the choice of initial value. We therefore say that the chain *asymptotically* have density $p(\cdot)$, after the burn-in period. We have set the burn-in to be the first 15.000 out of 40.000 total samples.[9]

Table 3.2: *Prior and posterior estimates from the Metropolis-Hastings algorithm. Non-inf. means non-informative.*

	Prior				**Posterior**			
	Dist.	mean	sd	initial	3%	50%	97%	sd
μ_0	Gaussian	0.0065	0.0001	0.0065	0.006277	0.006473	0.006645	0.000099
ρ	Gaussian	0.9700	0.001	0.97	0.968227	0.970080	0.971848	0.000986
$\rho_{r,\mu}$	Gaussian	-0.985	0.05	-0.955	-0.999875	-0.969047	-0.942929	0.017658
$\phi_{Expansion}$	Uniform	0.00	0.20	0.01	0.000180	0.003309	0.007802	0.002294
$\phi_{Recession}$	Uniform	0.00	0.20	0.145	0.135735	0.141191	0.147485	0.002913
h	Uniform	0.00	4.00	1.384	0.868431	1.311686	1.655376	0.228761
σ_0^2	Non-inf.	0.0015	0.00	0.0015	0.0015	0.0015	0.0015	0.00

Posterior Inference Results

Table 3.2 reports the percentiles of the posterior distribution. The narrower the credibility interval, the more certainty there is about the range where the true parameter value lies. We find

[9] M. Sugiyama (2016) [41]

all our posterior intervals to be very narrow. We find that the fraction of variance explained by the expected excess return, ϕ_{S_t}, greatly varies with the regime, as the posterior quantiles are significantly different. This indicates that expected excess returns are relatively stable during the expansionary regime, with $\phi_{Expansion}$ around 1%, and much more volatile during the recessionary regime, with $\phi_{Recession}$ around 15%. Hence, the variation in expected excess returns can be linked to movements in the business cycle. We further find that the two shock processes: surprises to excess return and to expected return, are very strongly negatively correlated ($\rho_{r,\mu} << 0$). This is in line with the findings of Campbell (1991) [23], who found that unexpected stock returns are negatively correlated with changes in expected future returns. Lastly, we find that the posterior median of the unconditional persistence of expected return, ρ, is 0.97, suggesting persistent excess return. This is in line with the findings in Gomez-Cram (2022) [14].

A simple way of summarizing the posterior distribution is to visualize it with a plot.[10] The posterior distributions of the model parameters can be seen in Figure 3.2 and belonging posterior trace can be seen in Figure 3.3. Most importantly, we find that the posterior distributions of the model parameters does not exhibit bimodal trends.

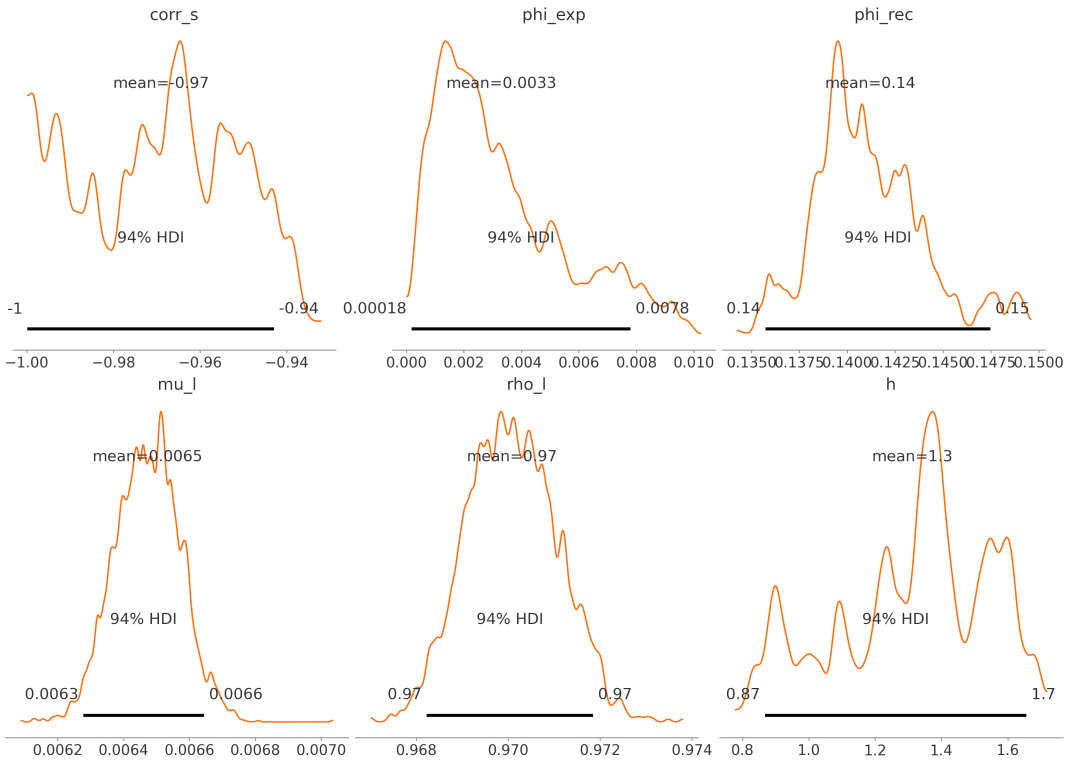


Figure 3.2: Posterior distributions *obtained from the Metropolis-Hastings algorithm of the model parameters. The sample period runs from January 1950 to August 2020.*

[10] We have applied the Python library arviz with functions 'plot_posterior' and 'plot_trace'.

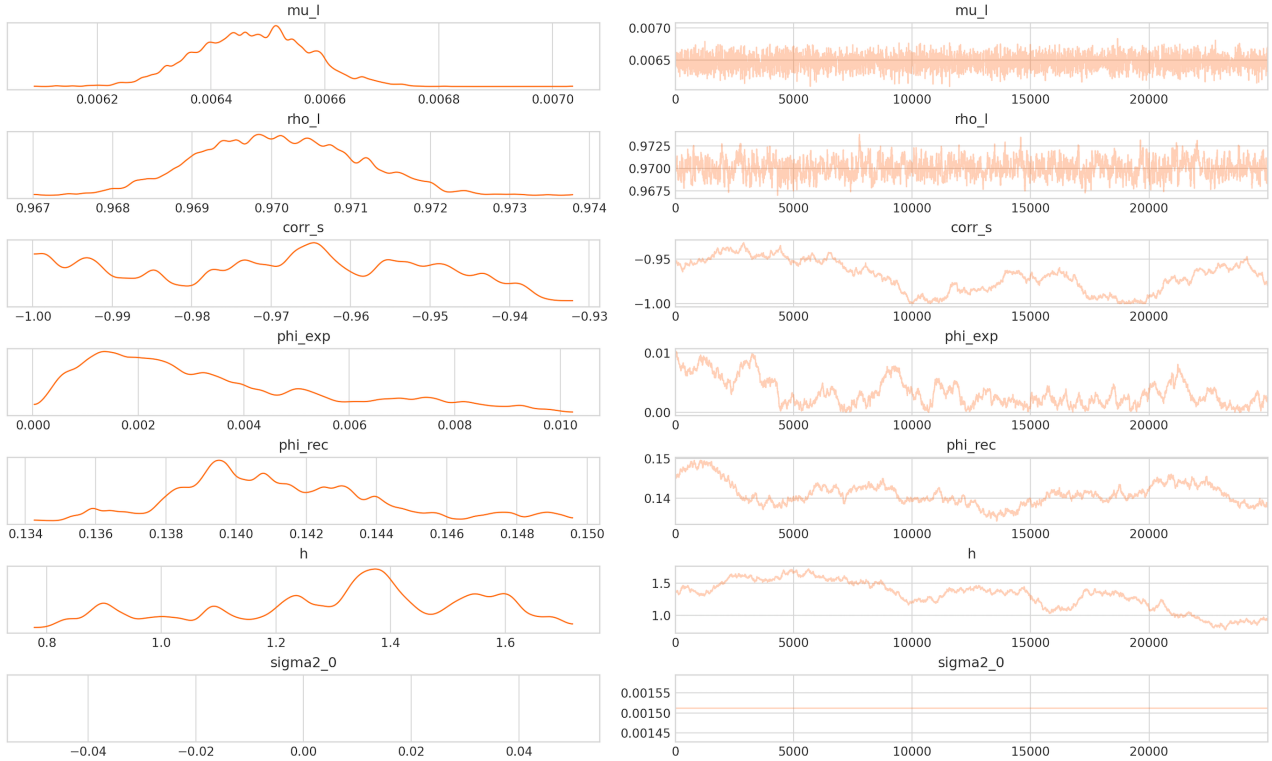


Figure 3.3: Posterior trace *of the Metropolis-Hastings algorithm. The sample period runs from January 1950 to December 2019.*

Posterior distribution and trace plots are great posterior inference tools, however not realistic for the time-varying variables. In the sample period January 1950 to December 2019 we would have 840 posterior distribution plots for the expected excess return alone. Instead the posterior median of every time period is plotted in Figure 3.4, to show the 1-month ahead forecast of excess return, $\hat{\mu}_{t+1|t}$, which was introduced in section 3.1.1.

From Figure 3.4 we see that the forecast of excess return varies across the unconditional/historical estimate. During the recessionary periods the estimate sharply drops to lower than average levels, whereafter it sharply rises again to higher than average levels. This trend is seen around all the recessions except the one beginning in 2001. We further find that negative forecasts are extremely rare, and, in the sample 1965-2022, such negative forecasts only really occurs twice, both in recessionary periods. Except during recessionary times, the estimate also drops below the unconditional average in the late 1990s, which is the period of the dot-com bubble.[11]

[11] *The value of equity markets grew exponentially during this period, with the technology-dominated Nasdaq index rising from under 1,000 to more than 5,000 between the years 1995 and 2000. Things started to change in 2000, and the bubble burst between 2001 and 2002 [...]* - [19]

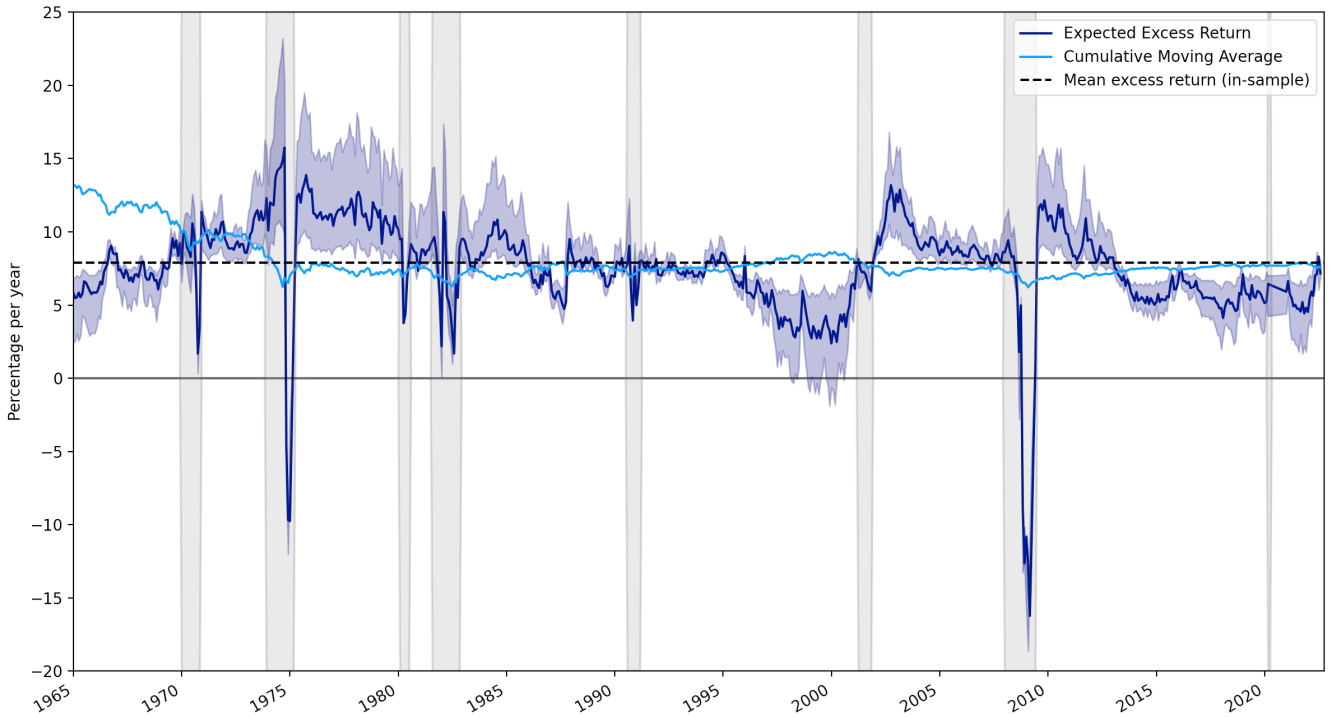


Figure 3.4: One month ahead forecast of excess return, $\hat{\mu}_{t+1|t}$. *The plot shows the posterior median estimate (navy), as well as the cumulative moving average of observed excess return (light blue) and the in-sample mean excess return of observed excess return (dotted black). The blue shaded area indicate 90% confidence bounds. The grey bars indicate NBER recession dates. The sample period runs from January 1965 to August 2022, with data from March 2020 to November 2020 removed.*

3.3 Regression-based forecast

In this section we perform a simple Ordinary Least Squares (OLS) regression of the excess return on our filtered probabilities to obtain a one month ahead forecast of excess return. The filtered probability of a recessionary regime was introduced in equation 2.9, derived in section 2.4 and shown in Figure 2.2. The regression of the excess returns on the filtered probability of recession, takes the following form

$$r^e_{t+1} = \alpha + \beta \cdot \hat{\pi}_{t|t} + \epsilon_{t+1} \tag{3.16}$$

Consider the recession probability as a deviation from an unconditional benchmark (sample mean)

$$r^e_{t+1} = \alpha + \beta \cdot (\hat{\pi}_{t|t} - \overline{\pi}) + \epsilon_{t+1} \tag{3.17}$$

We perform this regression and thereby obtain two forecasts of excess return

$$\begin{aligned} \mathbb{E}[r^e_{t+1}] &= \mu^{bmk} = \hat{\alpha} && \text{(Unconditional)} \\ \mathbb{E}[r^e_{t+1} \,|\mathcal{F}_t] &= \hat{\mu}^{reg}_{t+1|t} = \hat{\alpha} + \hat{\beta} \cdot (\hat{\pi}_{t|t} - \overline{\pi}) && \text{(Conditional)} \end{aligned}$$

We further perform lagged regression, meaning we would obtain the excess return h periods forward

$$r^e_{t+h} = \alpha + \beta^h \cdot (\hat{\pi}_{t|t} - \overline{\pi}) + \epsilon_{t+1} \tag{3.18}$$

$$\mathbb{E}[r^e_{t+h} \,|\mathcal{F}_t] = \hat{\alpha} + \hat{\beta}^h \cdot (\hat{\pi}_{t|t} - \overline{\pi}) \tag{3.19}$$

$$\Rightarrow \hat{\mu}^{reg,h}_{t+1|t} = \frac{1}{h} \cdot \Big(\hat{\alpha} + \hat{\beta}^h \cdot (\hat{\pi}_{t|t} - \overline{\pi})\Big) \tag{3.20}$$

The result of these regressions are shown in Table 3.3 and the $\hat{\beta}^h$ values are plotted in Figure 3.5.

Table 3.3: Regression result *of the regression of equation (3.18):* $r^e_{t+h} = \alpha + \beta^h \cdot (\hat{\pi}_{t|t} - \overline{\pi}) + \epsilon_{t+1}$. *The column numbers (1-12) indicate lags, i.e. values of h.*

	Dependent variable:											
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Intercept	0.073***	0.072***	0.072***	0.071***	0.072***	0.072***	0.071***	0.070***	0.069***	0.068***	0.067***	0.067***
	(0.026)	(0.019)	(0.016)	(0.014)	(0.012)	(0.011)	(0.010)	(0.010)	(0.009)	(0.009)	(0.008)	(0.008)
RecProb	-0.108	-0.112	-0.079	-0.051	-0.043	-0.025	-0.012	0.008	0.028	0.044	0.053	0.055
	(0.141)	(0.103)	(0.084)	(0.073)	(0.065)	(0.060)	(0.055)	(0.051)	(0.048)	(0.046)	(0.044)	(0.042)
Observations	419	417	415	413	411	409	407	405	403	401	399	397
R^2	0.001	0.003	0.002	0.001	0.001	0.000	0.000	0.000	0.001	0.002	0.004	0.004
Adjusted R^2	-0.001	0.000	-0.000	-0.001	-0.001	-0.002	-0.002	-0.002	-0.002	-0.000	0.001	0.002
Res. Std.	0.532	0.389	0.318	0.276	0.246	0.225	0.208	0.194	0.183	0.173	0.165	0.159
F Statistic	0.592	1.182	0.894	0.492	0.438	0.175	0.046	0.025	0.327	0.914	1.449	1.699
Note:												*p<0.1; **p<0.05; ***p<0.01

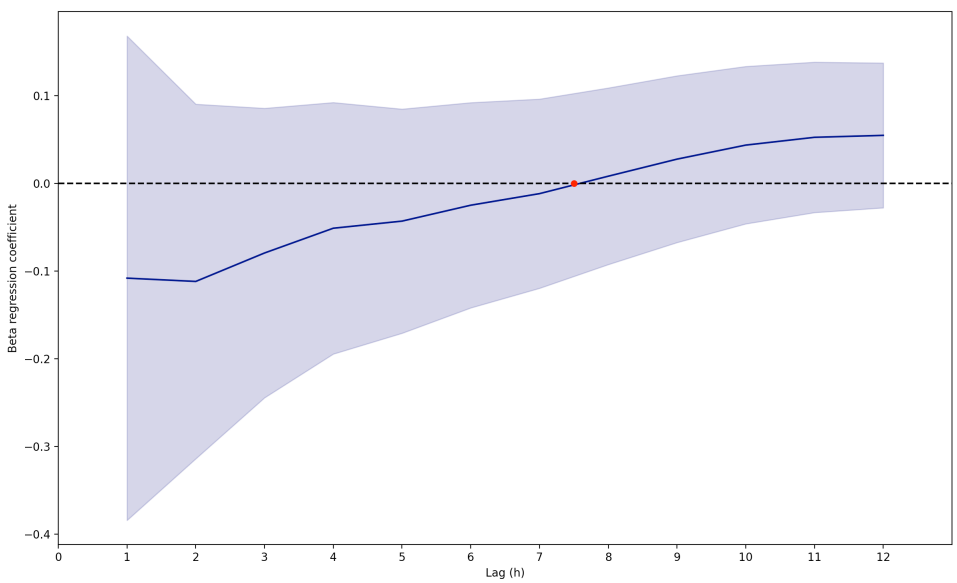


Figure 3.5: Betas. *The beta coefficients of the regressions of equation (3.18):* $r^e_{t+h} = \alpha + \beta^h \cdot (\hat{\pi}_{t|t} - \overline{\pi}) + \epsilon_{t+1}$. *The x-axis indicates number of lags in the regression (h value). The y-axis indicates the beta regression estimate,* $\hat{\beta}^h$. *The red dot is at* $(7.5, 0)$ *and indicates where the beta coefficient exceeds zero.*

From Figure 3.5 we find that only the lags up to and including 7 captures the expected negative recession coefficient. Hence, a too large lagged window, also catches the upturn of the recession and therefore the coefficient becomes positive. This indicates, that, when the recession probability exceeds it's average level, we predict there to be a positive equity premium to collect in approximately 8 months. We note that the confidence bounds are very wide (all include zero), and that

the beta coefficients are therefore not significant. We proceed with the analysis, however with our eyes wide open.

The forecast implications of the $\hat{\beta}^h$ estimate, can be seen in Figure 3.6, which plots the annualized expected excess return, of equation 3.20. We are interested in the 1 month lag, however also plot lag 6 and 12 for comparison. We choose these lags, as these are comparable to Figure 1.1 (the equity premium v-shape around recessions). Going forward we will refer to the lag 1 forecast, as the regression-based one month ahead forecast of excess return or simply the regression-based forecast. From Figure 3.6 we see that the regression-based forecast (blue line) varies with the business cycle. During expansionary periods the estimate is above the unconditional level (dotted black line). At the onset of recessions it sharply drops to, sometimes, negative levels, and thereafter sharply bounces back to be above the unconditional estimate, towards the end of the recessionary periods (the grey bars). We find that the effect is less pronounced in the 6 month lag and reverse in the 12 month lag. This is in line with the findings of Figure 1.1.

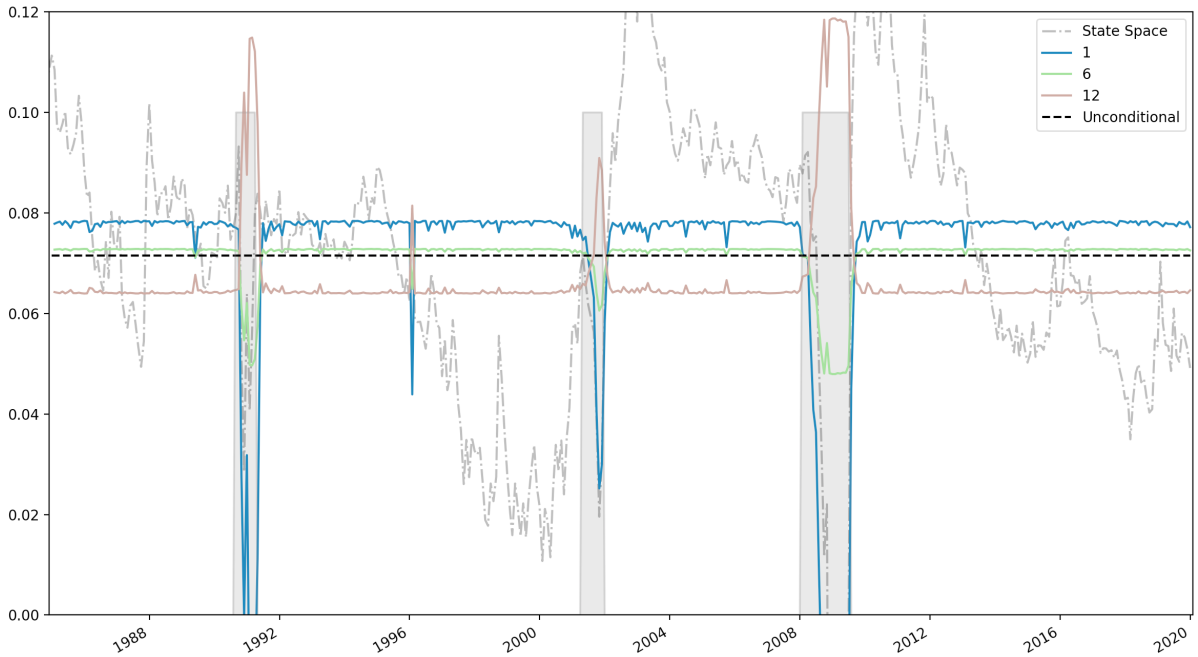


Figure 3.6: Equity forecast

Figure 3.7: Regression result. *The regression-based expected excess return forecast (colored lines) and, for comparison, the expected excess return forecasted from the state-space model (grey dotted) and unconditional (black dotted). The colored lines shows μ^{reg}, of equation 3.20: $\mu^{reg} = \frac{1}{h} \cdot \left(\hat{\alpha} + \hat{\beta}^h \cdot (\hat{\pi}_{t|t} - \overline{\pi})\right)$ with lag indicated by the legend. We have multiplied with 12 to show the annualized forecast of excess return. The sample period runs from January 1985 to December 2019. The grey bars show NBER recessions.*

Chapter 4

Timing the market

In this chapter we investigate the economic relevance and the ability to time the business cycle (i.e. catch economic instability) of our Chapter 2 and 3 estimated quantities: the filtered probability of recession, the common component and the one-month-ahead forecasts of excess return.

We start the chapter by investigating a simple market timing strategy, as Gomez-Cram (2022) [14], to test the timing ability of our estimates. In the next section, we expand the analysis to the mean-variance investor, following the methodology of Moench and Stein (2021) [30] and Moreira and Muir (2017) [31]. Here we examine two strategies, which employ the same filtered probabilities of recession, however exploit them differently - one directly in an Ordinary Least Square regression framework and the other one indirectly through a state space model. As a final note, we include a brief test of whether the results possibly expand to other factors, employing factor data of Pedersen et al. (2022) [42].

4.1 Simple Timing

In this section we investigate a simple market timing strategy, which switches from stocks to cash during bad economic times. The aim is to be fully invested in the stock market (S&P 500) when the macroeconomic times are good and be fully invested in cash (1-month Treasure Bill) during bad macroeconomic times.[1] We test and compare four different strategies, three with active switching conditions and one passive buy-and-hold for benchmark comparison. We apply a constant transaction cost of 2%.

Assume we are standing at time t, with the knowledge up to and including time $t-1$. We would like to infer whether we should invest in the stock market or hold cash.

The first strategy takes its position based only on the macroeconomic knowledge. More specifically, if the common component is below its moving average (MA), the strategy cashes out. The common component was introduced in equation (2.3) and deduced in section 2.4.1. The moving average window was determined by testing the strategy on a training time-frame, prior to

[1] By 'good' macroeconomic times we are referring to the expansionary regime and 'bad' refers to the recessionary regime.

initiation, in order not to create look-ahead bias.[2]

1. (Macro Timing) Common component [t-1] < MA [t-1] (20 months)

The second strategy takes its position based only on expected excess return. More specifically, if the 1-month ahead forecast of excess return is below its cumulative moving average (CMA), the strategy invests in T-bills. This component was introduced in equation (3.6) and derived in section 3.1.1. This component, however, is indirectly influenced by the business cycle.

2. (Return Timing) Forecast of excess return [t] < CMA [t-1]

The third strategy combines conditions 1 and 2.

3. (Combined Timing) Macro Timing AND Return Timing

The fourth strategy is a passive buy-and-hold strategy for benchmark comparison.

Figure 4.1 and Figure 4.2 plots the cumulative gains of the managed portfolios up against a benchmark buy-and-hold portfolio, for the sample periods 1965:1-2019:12 and 1985:1-2022:8 respectively. Here, we find that the combined strategy outperforms the passive benchmark on Sharpe ratios, with a 51.1% and 52.5% increase.

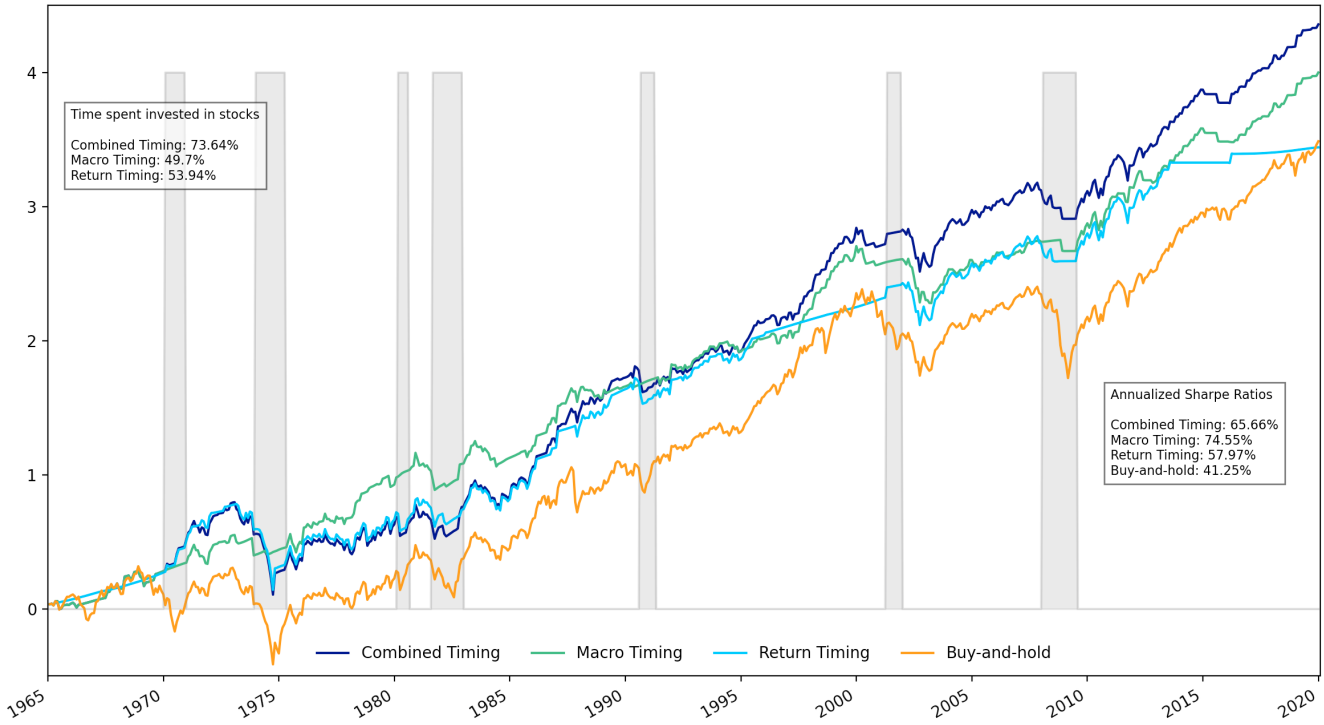


Figure 4.1: Cumulative returns *for the 3 simple timing strategies and the buy-and-hold benchmark. The sample period runs from January 1965 to December 2019.*

[2]We found, both in-sample and out-of-sample, that all window sizes, in a range from 6 to 36 months, outperformed the passive buy-and-hold.

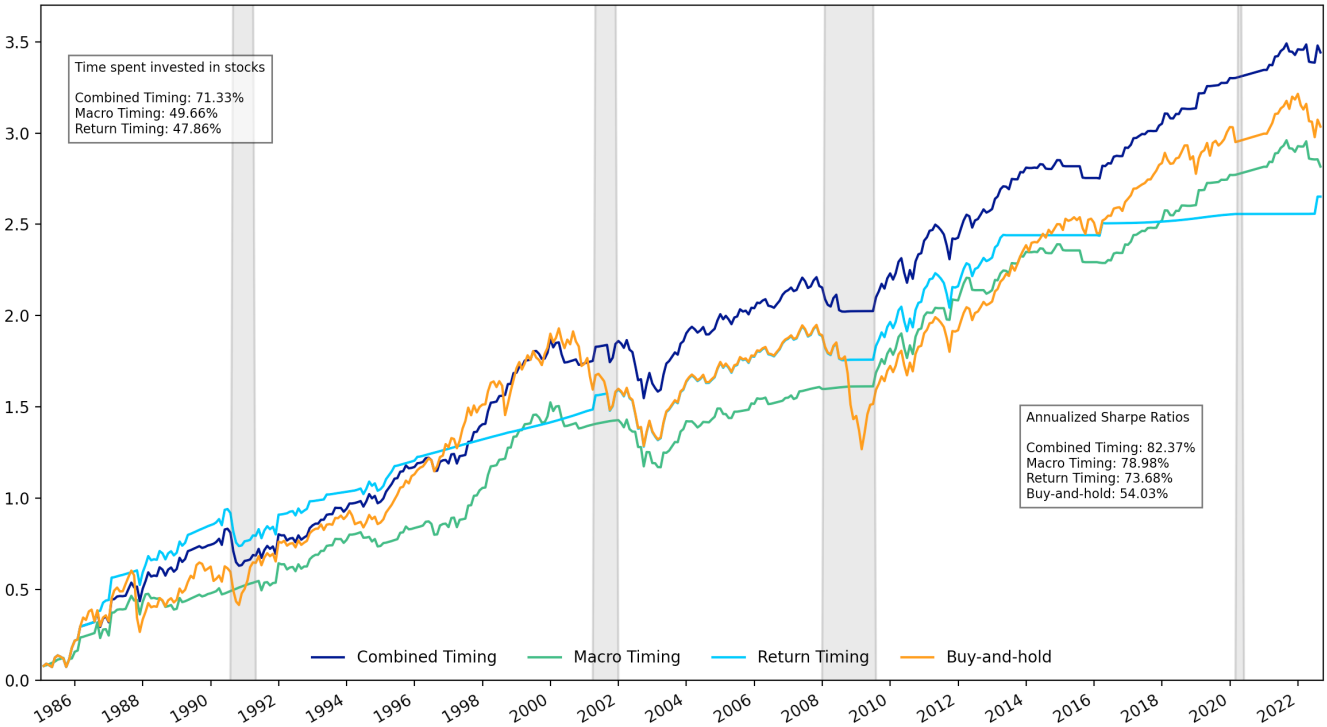


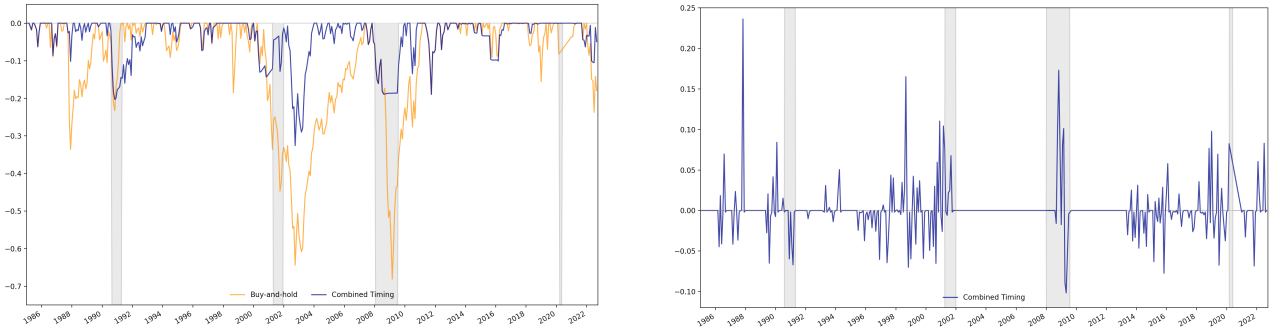
Figure 4.2: Cumulative returns *for the 3 simple timing strategies and the buy-and-hold benchmark. The sample period runs from January 1985 to December 2022, excluding March 2020 to November 2020.*

Over the longer sample period of Figure 4.1 we find that all three active timing strategies outperform the passive benchmark. The macro timing strategy (green) obtained the best annualized Sharpe Ratio with a significant increase of 70.9% over the passive benchmark.

In the shorter, and more recent, sample period of Figure 4.2, we find that the passive benchmark outperforms the market timing and the return timing in a cumulative way. We examine further to find that the timing strategies, which contains only one switching condition, also cashes out during expansionary periods. We note how especially the return timing (light blue) has several longer periods with a flat cumulative return during the expansionary regime. We find, however, that all three active timing strategies outperforms the passive on Sharpe ratios, with an increase of 46.2%, 36.4% and 52.5%, for strategy 1, 2 and 3 respectively. This could be due to the fact that the active strategies all manage to obtain a relatively flat cumulative return through the NBER recessions (grey bars), whereas the buy-and-hold strategy (orange) significantly drops around the recessionary periods.

To further examine when the combined timing strategy lose money relative to the buy-and-hold, Figure 4.3 plots the draw-downs and active returns. From Figure 4.3a, we find that the managed strategy benefits by reducing the maximum draw-down during recessions.[3] This is clearly seen as the blue curve is above the orange in Figure 4.3a.

[3]For an explanation of draw-down see [20]



(a) Draw-downs **(b) Active returns**

Figure 4.3: Return breakdown *of the combined timing strategy (navy) relative to the passive buy-and-hold benchmark (orange). The sample period runs from January 1985 to August 2022, excluding March 2020 to November 2020.*

The active return is the return in excess of the passive benchmark.

$$r \quad = \quad \underbrace{(r - r^{bmk})}_{\text{Active return}} \quad + \quad \underbrace{(r^{bmk})}_{\text{Passive return}}$$

An active return different from zero indicates that the timing strategy have switched from holding the market to holding cash. If the active return is positive, the strategy made the right bet, and reversely when negative. From Figure 4.3b, we find that there is a majority of correct bets, however also a high number of false positives during expansionary times. If we consider the areas around the grey NBER recession bars, we find the strategy successfully avoids the big recessionary down-leg. This is seen as the active return is either zero or positive leading up to, and during the first months of, the NBER recessions. However, we further see that the active timing strategy also misses out on some of the gains of the recessionary up-leg.[4] This is seen in the recessions beginning 1991 and 2008, where the active return is negative towards the end of the grey recession bars. Catching the recessionary up-leg is important, as Gomez-Cram (2022) [14] shows that a strategy, which switches to cash during the entire duration of the recessionary period, underperforms relative to a passive buy-and-hold strategy.

From Table 4.1 we find that all three of our timing strategies earns a significant positive CAPM-alpha, with the combined timing strategy earning the largest of 4.7% annualized.

[4]Recessionary down- and up-leg refers to the v-shape described in Introduction and illustrated in Figure 1.1.

Table 4.1: Performance of timing strategies. *Performance measures for 3 different timing strategies and a passive buy-and-hold benchmark. The timing strategies are fully invested in the stock market and switch to cash during bad economic times. How each strategy determines a 'bad economic time' is described in Section 4.1. The first column shows Sharpe ratios, with the parentheses showing the increase over the buy-and-hold benchmark. Alpha refers to the CAPM-alpha: the intercept of the time-series regression of equation (4.6) (in this column multiplied with 1200 to show annualized percentage). The parenthesis following indicates the p-value belonging to the alpha. Column 3 shows the* 90% *confidence bounds for the CAPM-alpha. The full regression result can be seen in Table 4.2. The sample period runs from January 1985 to August 2022, excluding March 2020 to November 2020.*

	Sharpe		**CAPM-alpha**		**Conf. bound**
Buy-and-hold	54.03	(0.00%)			
Market Timing	78.98	(46.2%)	4.325	(0.001)	(1.874, 6.775)
Return Timing	73.68	(36.4%)	3.832	(0.002)	(1.367, 6.298)
Combined Timing	82.37	(52.5%)	4.747	(0.000)	(2.286, 7.208)

Table 4.2: Regression result *of the time series regression test of the managed portfolio on the non-managed market factor:* $r_{t+1}^{Timing} = \alpha + \beta \cdot r_{t+1}^{e} + \epsilon_{t+1}$, *where* r_{t+1}^{e} *denotes the market excess return.*

	Dependent variable:		
	Market	Return	Combined
Intercept	0.004***	0.003***	0.004***
	(0.001)	(0.001)	(0.001)
r^e	0.033***	0.034***	0.046***
	(0.002)	(0.002)	(0.002)
Observations	443	443	443
R^2	0.401	0.405	0.560
Adjusted R^2	0.400	0.403	0.559
Resid. Std. Error	0.022	0.022	0.022
F Statistic	295.147***	299.782***	561.576***
Note:	*p<0.1; **p<0.05; ***p<0.01 and (df = 441)		

Transaction cost

Throughout this section, we have made sure to account for transaction costs, when switching from one financial instrument to another. We have applied constant transaction costs of 2%. As we argued in the introduction, this thesis is not for the common small investor, which is why an investor, who is able to employ the algorithm in this thesis, might make transactions so large that it would shift the market with an all or nothing strategy like this.

The informed investor is trading and moving prices because he has some knowledge that others do not have. But when these trades are large, they move prices - A. Ang (2014) [1]

Chen, Lezmi, Roncalli and Xu (2019) [34] discusses how unit transaction costs may be a linear function of the trading size. They further introduce quadratic transaction costs in the mean-variance portfolio optimization. Gârleanu and Pedersen (2014) [11] covers an analysis of optimal dynamic portfolio policy when trading is costly, and further assume, that transaction costs are proportional to the amount of risk. A possible extension of this thesis could be to test if the results hold under different transaction cost assumptions.

4.2 Dynamic Timing

In this section we consider a dynamic strategy where the timing of the business cycle is incorporated in the equity premium forecast. We consider a mean-variance investor who allocates funds across a risky and risk-free asset. At time t, the investor optimally decides to invest the following share in the risky asset,

$$w_t^c = \frac{1}{\gamma} \cdot \frac{\hat{r}_{t+1|t}}{\hat{\sigma}_t^2} \tag{4.1}$$

Here γ is the relative risk aversion and $\hat{r}_{t+1|t}$ is the 1-month ahead forecast of excess return. We employ two different forecasts:

1 The state-space forecast, $\hat{\mu}_{t+1|t}$, as introduced in (3.6) and derived in section 3.1.1.

2 The regression-based forecast, $\hat{\mu}_{t+1|t}^{reg}$, as introduced and derived in section 3.3.

The state-space forecast utilizes the information contained in the filtered probability of recession, indirectly through a state-space model.[5] The regression-based forecast utilizes the filtered probability of recession, directly through an OLS regression. To facilitate model interpretation we let $\hat{\sigma}_t^2$ be the unconditional variance, thus the weights in the risky asset only differ due to the equity premium forecasted by the given model. The superscript c stands for conditional, thus indicating the strategy employs the conditional estimate of excess return. The excess realized portfolio return at time t+1 is

$$r_{t+1}^c = w_t^c \cdot r_{t+1}^e \tag{4.2}$$

[5]More specifically the filtered probability of recession is a component in the predictive density of the expected excess return.

Where r^e_{t+1} is the excess market return. To more directly examine the economic gain of the conditional estimates relative to the unconditional benchmark, we consider a strategy which takes a long position in the conditional estimate and short in the unconditional estimate, i.e. the long-short portfolio weight is

$$w^{ls}_t = \frac{1}{\gamma} \cdot \left(\frac{\hat{r}_{t+1|t}}{\hat{\sigma}^2_t} - \frac{\hat{r}_t}{\hat{\sigma}^2_t} \right) \tag{4.3}$$

where

$$\begin{aligned} \hat{r}_{t+1|t} &= \mathbb{E}[r^e_{t+1} \,|\mathcal{F}_t] \quad \text{(Conditional)} \\ \hat{r}_{t+1} &= \mathbb{E}[r^e_{t+1}] \qquad \text{(Unconditional)} \end{aligned}$$

We rebalance at the beginning of each month, after new macroeconomic data have been released. The estimates used in the strategies are determined on real-time data. This means 1) we only employ the not-revised macroeconomic data that were available to investors at the time and 2) for the historical estimates the sample size increases with each month. We start the strategy and the out of sample period in January 1985, with data from January 1950 up to January 1985. The risk aversion parameter is assumed constant and chosen such that the unconditional portfolio weight is equal to one on average.[6] Campbell and Cochcrane (1999) [3] show that the common investors risk aversion varies with the business cycle, thus indicating a risk premium if we keep constant risk aversion across regimes.

The performance of the long-short strategy is an indication of the economical relevance of the conditional estimate. Positive portfolio returns, would indicate great economic relevance in applying the conditional estimate as opposed to the unconditional.

We will, amongst other, evaluate our result as done in Moench and Stein (2021) [30] who uses two measures: the certainty equivalent return (CER) and Sharpe ratio (SR). As stated in Chen et al., 2020 [22], the CER multiplied by 12 can be interpreted as the annual risk-free rate that an investor is willing to accept instead of holding the risky portfolio. By comparing two CERs we obtain a quantity which can be interpreted as the annual portfolio management fee that an investor would be willing to pay to have access to the alternative forecasting model.[7]

$$\text{CER} = \hat{\mu}_P - \frac{\gamma}{2} \cdot \hat{\sigma}^2_P \tag{4.4}$$

$$\text{Sharpe} = \frac{\hat{\mu}_P \cdot 12}{\hat{\sigma}_P \cdot \sqrt{12}} \cdot 100 \tag{4.5}$$

Where $\hat{\mu}_P$ and $\hat{\sigma}^2_P$ are the mean and variance of the realized portfolio returns. We multiply the Sharpe ratio with $\frac{12}{\sqrt{12}}$, to obtain the annualized Sharpe ratio, and further with 100 to obtain the result in percentage.

To further test the economic relevance of our results we follow the empirical methodology of

[6] For the sample 1985:1-2019:12, the risk aversion parameter is approximately 3.5.

[7] This is dependent on the investors risk preferences and leverage constraints. Moench and Stein (2021) [30] test different risk aversion and leverage constraints and find that the gains are highest when the risk aversion is 3 and leveraging and short-selling up to 50% is allowed.

Moreira and Muir (2017) [31] and include a time-series regression test of the managed market factor on the non-managed market factor. This factor can be easily traded in the market. [8]

$$r^c_{t+1} = \alpha + \beta \cdot r^e_{t+1} + \epsilon_{t+1} \tag{4.6}$$

A positive alpha in this regression would imply that the managed factor increases the Sharpe ratio relative to the non-managed market factor. As further argued by Moreira and Muir (2017) [31], a positive alpha implies that the managed strategy expands the mean-variance frontier. If we calculate the root mean squared error (RMSE), we can construct the appraisal ratio (AR) of the managed portfolio. AR measures how the slope of the mean-variance efficient frontier, spanned by the original market factor, is expanded.[9]

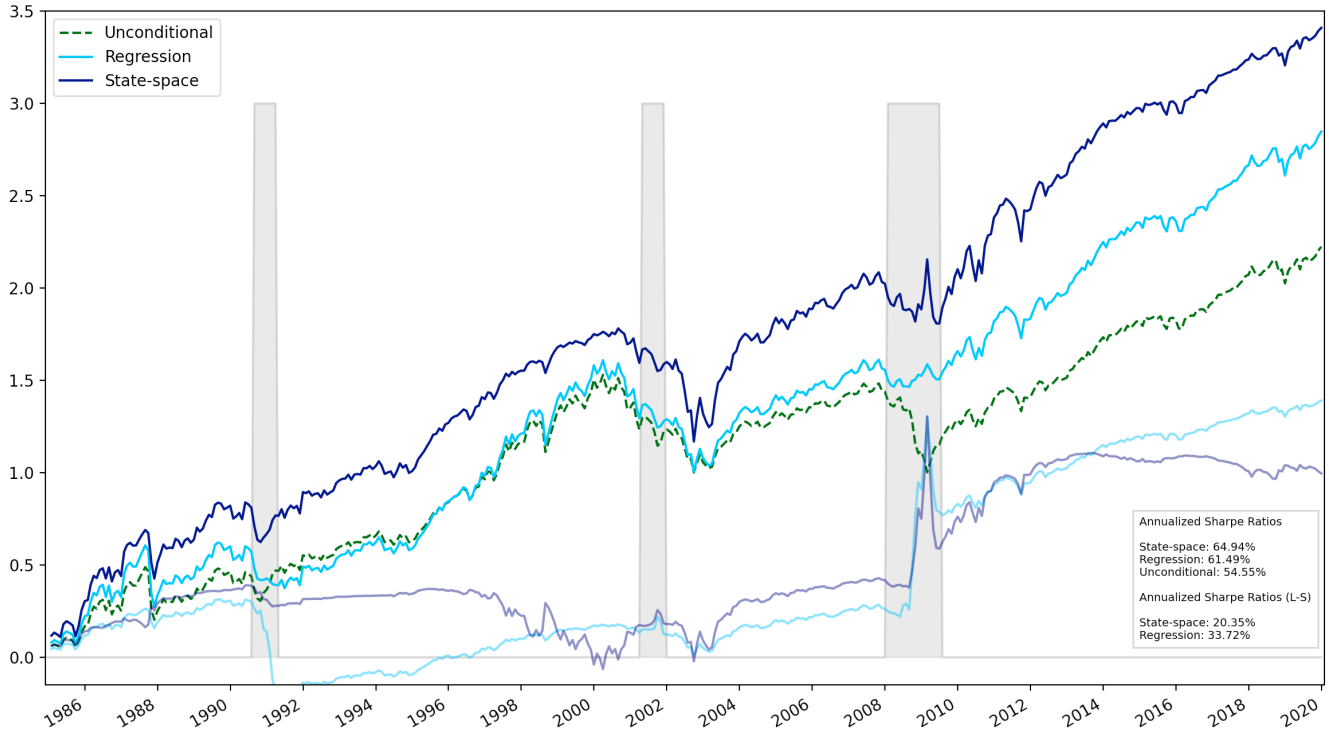


Figure 4.4: Cumulative returns *for the conditional mean-variance strategies (dark and light blue) as compared to the unconditional (dotted green), as well as the cumulative gains for the long-short strategies (faint blues). for the 6 regression based mean-variance strategies (pastel colors) as compared to the market (red) and the strategy employing the state-space equity forecast (navy). The sample period runs from January 1985 to December 2019. The grey bars show NBER recessions.*

From Figure 4.4 we find that both strategies which employs the conditional estimates outperform the historical benchmark (green dotted line). This fact is seen as the conditional strategies cumulatively ends much above the unconditional, earns larger Sharpe ratios, and further, as the long-short strategies (faint blues) cumulatively ends well above zero with positive Sharpe ratios.

Comparing the two conditional strategies, we see that the cumulative returns of the strategy employing the state-space forecast, $\hat{\mu}_{t+1|t}$, is consistently above the strategy employing the regression-based forecast, $\hat{\mu}^{reg}_{t+1|t}$. This point, however, does not transfer to the long-short cumulative returns.

[8] This could for example be an index tracking the S&P500.

[9] For a recap of RMSE see [38].

Here the long-short strategy employing the regression-based estimate outperforms the state-space and earns a larger terminal cumulative return, as well as a 64% increase in Sharpe ratio. We find that, especially during the recession beginning in 2008, both long-short strategies experienced large positive gains, however also followed by a large drop in cumulative gains.

To further examine when our strategy loses money relative to the unconditional, we examine the portfolio weights, active returns, and the draw-downs of the conditional strategies relative to the unconditional. This return analysis is shown in Figure 4.5.

Figure 4.5b plots the draw-downs of the conditional mean-variance strategy based on the state-space forecast. We find that both the conditional and unconditional strategy suffers draw-downs, with the largest occurring around recessionary periods, however the conditional strategy have reduced maximum draw-downs during recessions, compared to the unconditional. This is seen as the blue curve is below the green. Further the conditional strategy quickly bounces back, whereas the unconditional strategy takes longer to recoup the recessionary losses. Figure 4.5a plots the draw-downs of the conditional mean-variance strategy based on the regression-based forecast. Here the same trend only shows around the recession starting in 2008. During the recession starting in 1991 we even find that the conditional strategy suffers greater draw-downs than the unconditional.

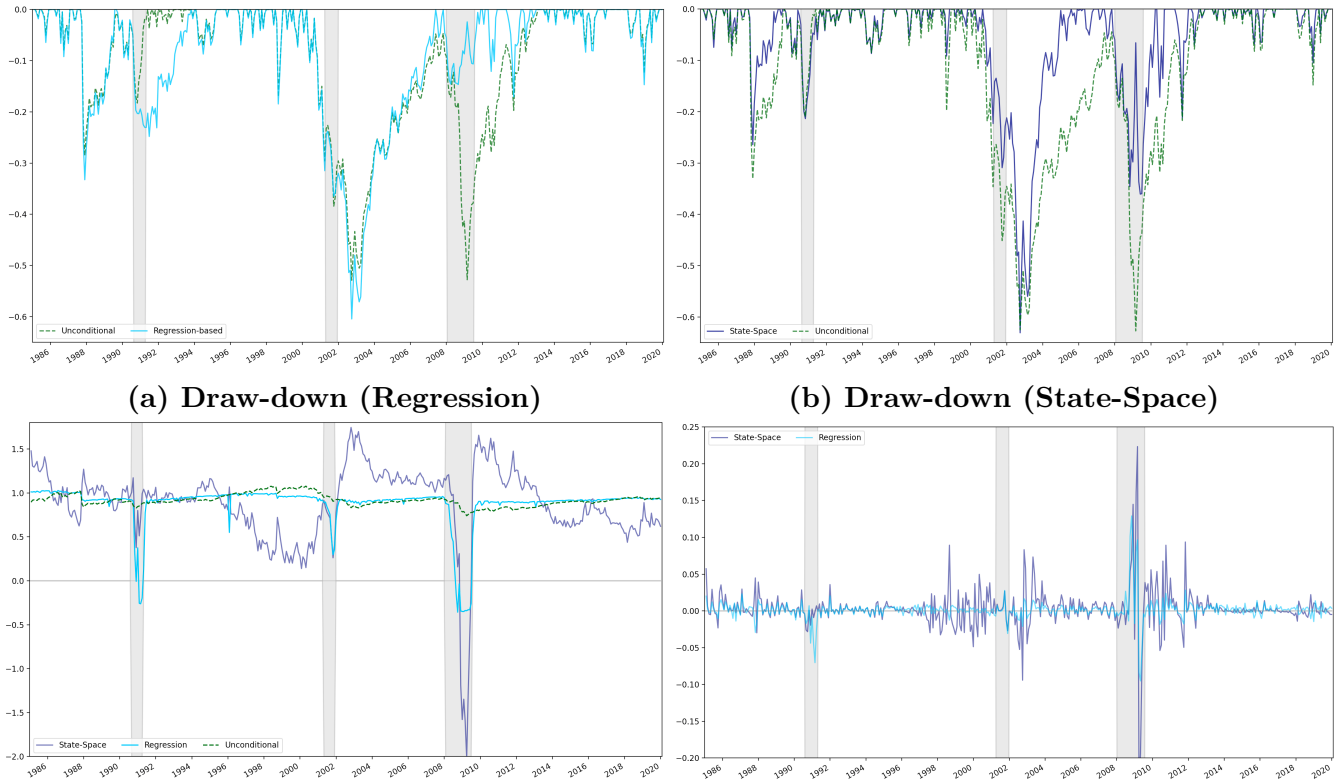
Figure 4.5c plots the portfolio weights of the three mean-variance strategies. We, expectedly, find the unconditional portfolio weights, based on the historical average, to be continuously positive, non-varying and approximately equal to one.

From Figure 4.5c we find that the regression-based strategy takes a position approximately equal to the unconditional during expansionary periods. At the onset of the recessions starting in 1991 and 2008 the portfolio weights sharply drops to negative levels, and thereafter bounces back to again be at the same level as the unconditional, towards the end of the recessionary periods (the grey bars). This means that the conditional strategy will carry the same risk as the unconditional during expansionary periods and less risk during recessionary periods, and further even shorting risk, during the recessions beginning in 1991 and 2008. This is seen as the portfolio weight in the risky asset drops or becomes negative during recessionary times. As seen in Figure 4.5d, this results in either very small or zero negative active returns during expansionary times, as well as positive active returns leading up to and during the beginning of the recession. We find, however, that the regression-based strategy does not bounce back fast enough, and thus misses out on some of the gains of the recessionary up-leg. This is seen as the active return is negative towards the end of the grey recession bars.

From Figure 4.5c we find that the portfolio weights of the state-space-based conditional strategy vary a lot. A pattern is clearly seen around the recession beginning in 2008, where the portfolio weights are clearly above 1 leading up to the recession, then, as the recessionary regime starts, sharply drops to become negative, and then quickly reverts back to more than 1 towards the end of the recession. This is the same period the long-short strategy (faint navy in Figure 4.4) experiences very significant gains. In this period, the strategy cashes out in the recession (negative portfolio weights), whereafter it significantly increases risk-taking (higher than average portfolio weights).

In conclusion, the conditional strategy based on the state-space forecast, benefits by reducing

the maximum draw-down during recessions and by returning more powerful to the market towards the end of a recession, and thereby more quickly bounce back. This is achieved as, the strategy takes on less risk during recessions (small or negative portfolio weights), whereafter it significantly increases risk-taking (higher than average portfolio weights). The regression-based conditional strategy benefits by greatly reducing the 2008 recession draw-down and further by not taking a contractionary position of the unconditional during expansionary periods.



(a) Draw-down (Regression) **(b) Draw-down (State-Space)**

(c) Portfolio weights **(d) Active returns**

Figure 4.5: Return analysis *of the conditional mean-variance strategies based on the State-Space forecast (navy), the regression-based (light blue) and the unconditional historical (green dotted). Panel a and b shows the draw-downs of the conditional strategies plotted up against the unconditional. Panel c plots the dynamic portfolio weights of the three strategies. Panel d shows the active returns. The passive benchmark, which the active strategies are measured against, is the unconditional, which applies the historical average as expected excess return. The sample period runs from January 1985 to December 2020.*

Table 4.3 reports the dynamic strategies' performance results. We find that the dynamic mean-variance strategy, which employs the unconditional historical estimate earns a CER of 7.15 and Sharpe ratio of 54.55%.

Examining the performance of the dynamic mean-variance strategy, which employs the simple OLS regression-based forecast of excess return, we find this to significantly outperform the unconditional. The regression-based conditional strategy earns a Sharpe ratio of 61.49% (12.7% increase), CAPM-α of 1.33% annualized (p-value of 0.19), annualized appraisal ratio of 0.23 and further a CER of 8.06 (0.91 difference), the latter means an investor would be willing to pay 91 bps annually to have access to the simple OLS forecast. We further find there to be statistically

significant economic relevance in our filtered recession probabilities, as the long-short strategy earns a Sharpe ratio of 33.72%, as well as a statistically significant (on a 5% level) result of a t-test on whether the long-short portfolio returns are positive.

Examining the performance of the dynamic mean-variance strategy, which employs the state-space forecast of excess return, we find this to also outperform the unconditional. The state-space based conditional strategy earns a Sharpe ratio of 64.83%, which is a 18.9% increase over the unconditional and a 5% increase over the regression-based, as well as an appraisal ratio of 0.33, which is an increase of 43.5% over the regression-based. The state-space based strategy further increases CER earning 8.71, an 8% increase over the regression-based strategy and an increase of 1.56 points over the unconditional, indicating that an investor would be willing to pay an annual portfolio management fee of 156 bps to have access to the forecasting model. We further find there to be statistically significantly economic relevance in our state-space based one month ahead forecast of excess return, as the managed portfolio earns a statistically significant (on a 5% level) CAPM-α of 3.37% annualized and beta of 0.6 (on a 1% level). We find this result to be comparable to the results of Moreira and Muir (2017) [31] who found that their volatility scaled market portfolio earned an annualized alpha of 4.86% and a beta of 0.6.[10] As done in the analysis of Moreira and Muir (2017) [31] we can calculate the percentage utility gain, with an annualized appraisal ratio of 0.33 and a market buy-and-hold Sharpe ratio of 57.33%, we find

$$\text{Utility Gain (\%)} = \frac{\text{SR}^2_{new} - \text{SR}^2_{old}}{\text{SR}^2_{old}} = \frac{\sqrt{0.5733^2 + (\frac{0.33}{\sqrt{12}})^2} - 0.5733^2}{0.5733^2} = 0.76820 \tag{4.7}$$

This implies that a mean-variance investor, who can only trade the market portfolio, can increase her lifetime utility by 76.8%, by applying our state-space forecast. The old Sharpe ratio is that of the buy-and-hold market portfolio, hence the reader must take into consideration that this result is obtained without accounting for e.g. transaction cost of the managed portfolio.

In regards to the difference in significance levels of the t-test (positive long-short returns) and the CAPM-α, we suspect this to be due to the position taking during expansionary periods, where the regression-based strategy stays close to the unconditional level, whereas the state-space based portfolio weights varies more around the unconditional level. We found that this variation creates negative active returns also during expansionary periods, and thus a less significant t-test of positive long-short returns. For the regression-based strategy this position would mean returns would not differ much from the market, thus the significance level of the difference to the non-managed market factor decreases, i.e. the CAPM-α.

[10] For the result see Moreira and Muir (2017) [31] Table I on page 1619.

Table 4.3: Dynamic Strategy Performance Results. *Portfolio performance result for the conditional strategies employing the OLS and State-Space equity premium forecast, as well as the unconditional benchmark employing historical estimates. The first column reports the certainty equivalent return (CER), as defined in equation 4.4. The second column reports Sharpe ratios, as defined in equation 4.5. The parentheses in these two columns shows the increase over the unconditional strategy. The third column reports the Sharpe ratio of the Long-Short portfolio, as defined in equation 4.3. The fourth column,* $\mu_{LS} > 0$*, reports the p-value of the t-test on whether the long-short portfolio returns are positive. The fifth column, CAPM-*α*, reports the intercept of the time-series regression of equation (4.6) (multiplied with 1200 to show annualized percentage). The parenthesis following indicates the p-value belonging to this estimate. AR is an abbreviation for appraisal ratio, and is given by* $\sqrt{12}\frac{\alpha}{RMSE}$*. The RMSE is also a result of the time-series regression. The full regression result is reported in Table 4.4. The sample period, for all reported results, runs from January 1985 to December 2019.*

	CER		**Sharpe**		**L-S**	$\mu_{LS} > 0$	CAPM-α		**AR**
Unconditional	7.15	(0.00)	0.55	(0.00%)					
Regression	8.06	(0.91)	0.61	(12.7%)	0.34	0.047	1.33	(0.19)	0.23
State-Space	8.71	(1.56)	0.65	(18.9%)	0.21	0.226	3.37	(0.05)	0.33

Table 4.4: Regression result *of the time-series regression of the managed market factor on the non-managed market factor, equation (4.6):* $r^c_{t+1} = \alpha + \beta \cdot r^e_{t+1} + \epsilon_{t+1}$

	Dependent variable:	
	State-Space	Regression
Intercept	0.003*	0.001
	(0.001)	(0.001)
Market	0.062***	0.065***
	(0.003)	(0.002)
Observations	420	420
R^2	0.550	0.804
Adjusted R^2	0.549	0.803
Residual Std. Error	0.029(df = 418)	0.017(df = 418)
F Statistic	511.145*** (df = 1.0; 418.0)	1709.520*** (df = 1.0; 418.0)
Note:		*p<0.1; **p<0.05; ***p<0.01

4.3 Future research

4.3.1 Factor Investing

We showed in section 4.1 that by timing the stock market we obtain a return in excess of the passive buy-and-hold benchmark. The allocation between equity and bond has been shown to be the most important decision taken, hence it is the primary influence on return. Brinson, Hood, and Beebower (1986) [12] show that approximately 90% of the variance of a typical funds return is explained by the asset allocation decision, i.e. the allocation between equity and bonds. Ang, Goetzmann, and Schaefer (2009) [2] did a large study of the Norwegian Government Pension Fund managed by Norges Bank Investment Management (NBIM) and, amongst other, found that 99.1% of variation in the funds returns, in the period January 1998 to September 2009, was explained by the equity-bond decision taken by Parliament. Ang et al. (2009) [2] conclude and recommend the Norwegian government to continue to employ active management, however, to focus on dynamic factor investing instead of the equity/bond allocation.

In section 4.2 we find that a simple regression-based managed market factor outperforms the unmanaged market factor. In this section we include a brief test of whether this expands to other factors. A whole thesis could be written on factor investing, however this is not the primary focus of this thesis. We assume the factors can be traded in the market and employ data from Pedersen et. al. (2022) [42]. There are 13 different factor themes and 153 specific factors.

We follow the previous section and perform an in-sample regression of each factor theme on the time t recession probability as a deviation from an unconditional benchmark (sample mean)

$$f^i_{t+1} = \alpha^i_1 + \beta^i_1 \cdot (\hat{\pi}_{t|t} - \overline{\pi}) + \epsilon_{t+1}, \quad i \in \{\text{factors themes}\} \tag{4.8}$$

The result of this regression is in Table 4.6 and Table 4.5, under Regression 1. It is especially noteworthy to examine the p-values. These utilizes the t-statistic to produce a measure of the likelihood of the β_1 coefficient. For example we find that 'Profitability' obtains a p-value of 0.970, meaning there is a 97.0% chance that the predictor, $(\hat{\pi}_{t|t} - \overline{\pi})$, has no effect on the dependent variable, 'Profitability'. By expecting the p-values we find that most of the factor themes have β_1 values insignificant from zero, with the insignificant p-values ranging from 0.970 down to 0.243. Out of 13 themes, we find that 4 factor themes produces significant p-values (these are highlighted with bold in Table 4.5). Next, we employ the estimates to compute the managed portfolios, F^i_{t+1}, for $i \in \{\text{factors themes}\}$,

$$\mathbb{E}[f^i_{t+1} \,|\, \mathcal{F}_t] = \hat{\alpha}^i_1 + \hat{\beta}^i_1 \cdot (\hat{\pi}_{t|t} - \overline{\pi}) \quad \Rightarrow \quad F^i_{t+1} = \mathbb{E}[f^i_{t+1} \,|\, \mathcal{F}_t] \cdot f^i_{t+1} \tag{4.9}$$

We, again, follow the methodology of Moreira and Muir (2017) [31] and perform a time-series regression test of the managed factor, F^i_{t+1}, on the non-managed factor, f^i_{t+1}.

$$F^i_{t+1} = \alpha^i_2 + \beta^i_2 \cdot f^i_{t+1} + \epsilon_{t+1}, \quad i \in \{\text{factors themes}\} \tag{4.10}$$

The result is stated in Table 4.7 and Table 4.5, under Regression 2. We find that all the managed factors outperforms the non-managed counterpart. The only exception is 'low leverage', which is inconclusive. We find it odd that all these results are significant, when the belonging $\hat{\beta}_1$ coefficient was not. The factors, which have a statistically significant $\hat{\beta}_1$ coefficient, are Momentum, Profit growth, Short term reversal and Size. These earn annualized alpha's of 3.73%, 1.39%, 1.03% and

0.3% respectively. The first is comparable to the results of Moreira and Muir (2017) [31] who found that their managed market portfolio earned an annualized alpha of 4.86%.

Table 4.5: Result of the in-sample factor regression. *Regression 1 refers to the regression described in Equation (4.8):* $f^i_{t+1} = \alpha^i_1 + \beta^i_1 \cdot (\hat{\pi}_{t|t} - \overline{\pi}) + \epsilon_{t+1}$. *Regression 2 refers to the regression described in Equation (4.10):* $\mathbb{E}[f^i_{t+1} \,|\, \mathcal{F}_t] \cdot f^i_{t+1} = \alpha^i_2 + \beta^i_2 \cdot f^i_{t+1} + \epsilon_{t+1}$. *We have multiplied the intercept of this regression,* $\hat{\alpha}_2$, *with 1200 to show annualized percentage. The sample period runs from January 1985 to December 2021.*

	Regression 1			**Regression 2**		
	$\hat{\beta}^i_1$	**Conf. bound**	**p-value**	$\hat{\alpha}^i_2$	**Conf. bound**	**p-value**
Accruals	-0.0018	(-0.008, 0.004)	0.567	2.0222	(1.989, 2.056)	0.000
Debt issuance	0.0014	(-0.003, 0.006)	0.537	2.5140	(2.487, 2.541)	0.000
Investment	0.0014	(-0.013, 0.016)	0.838	3.1572	(3.131, 3.184)	0.00
Low leverage	0.0072	(-0.014, 0.028)	0.502	-0.0646	(-0.195, 0.066)	0.330
Low risk	-0.0073	(-0.032, 0.017)	0.556	2.2759	(2.144, 2.408)	0.000
Momentum	**-0.0334**	(-0.054, -0.013)	**0.001**	3.7307	(3.135, 4.327)	0.000
Profit growth	**-0.0048**	(-0.011, 0.002)	**0.153**	1.3915	(1.305, 1.478)	0.000
Profitability	-0.0002	(-0.015, 0.015)	0.970	3.9676	(3.963, 3.973)	0.000
Quality	0.0054	(-0.004, 0.015)	0.243	3.7610	(3.662, 3.860)	0.000
Seasonality	0.0021	(-0.001, 0.006)	0.252	1.5140	(1.476, 1.552)	0.000
Short term rev.	**-0.0097**	(-0.017, -0.002)	**0.014**	1.0269	(0.853, 1.201)	0.000
Size	**0.0130**	(0.000, 0.026)	**0.048**	0.2955	(0.062, 0.529)	0.013
Value	0.0068	(-0.015, 0.029)	0.545	3.8304	(3.708, 3.953)	0.000

	Dependent variable:						
	accruals	debt issuance	investment	low leverage	low risk	momentum	profit growth
Intercept	0.002***	0.002***	0.003**	-0.000	0.002	0.003*	0.001**
	(0.001)	(0.000)	(0.001)	(0.002)	(0.002)	(0.002)	(0.001)
RecessionProb	-0.001	0.001	0.000	0.009	-0.013	-0.037***	-0.006*
	(0.003)	(0.002)	(0.007)	(0.011)	(0.013)	(0.010)	(0.003)
Observations	435	435	435	435	435	435	435
R^2	0.000	0.000	0.000	0.001	0.002	0.029	0.009
Adjusted R^2	-0.002	-0.002	-0.002	-0.001	0.000	0.026	0.006
Resid. Std. Error	0.011	0.008	0.024	0.036	0.042	0.034	0.011
F Statistic	0.053	0.047	0.000	0.636	1.037	12.799***	3.731*

	Dependent variable:					
	profitability	quality	seasonality	short term reversal	size	value
Intercept	0.003***	0.003***	0.001***	0.001	0.000	0.003*
	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.002)
RecessionProb	-0.003	0.001	0.000	-0.009**	0.018***	0.007
	(0.007)	(0.005)	(0.002)	(0.004)	(0.007)	(0.011)
Observations	435	435	435	435	435	435
R^2	0.000	0.000	0.000	0.011	0.017	0.001
Adjusted R^2	-0.002	-0.002	-0.002	0.009	0.015	-0.001
Resid. Std. Error	0.024	0.015	0.006	0.013	0.022	0.037
F Statistic	0.174	0.093	0.000	4.801**	7.571***	0.446
Note:						*p<0.1; **p<0.05; ***p<0.01 and (df = 433)

Table 4.6: Regression 1 result. The result of the regression described in Equation (4.8): $f^i_{t+1} = \alpha^i_1 + \beta^i_1 \cdot (\hat{\pi}_{t|t} - \overline{\pi}) + \epsilon_{t+1}$.

	Dependent variable:						
	accruals	debt issuance	investment	low leverage	low risk	momentum	profit growth
Intercept	0.002***	0.002***	0.003***	-0.000	0.002***	0.003***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
r_e	0.001	0.001	0.000	0.001	0.001	0.023***	0.005
	(0.001)	(0.001)	(0.000)	(0.002)	(0.001)	(0.007)	(0.003)
Observations	434	434	434	434	434	434	434
R^2	0.001	0.001	0.000	0.001	0.001	0.023	0.005
Adjusted R^2	-0.002	-0.001	-0.002	-0.001	-0.002	0.021	0.002
Residual Std. Error	0.000	0.000	0.000	0.001	0.001	0.005	0.001
F Statistic	0.327	0.382	0.042	0.451	0.341	10.269***	2.045

	Dependent variable:					
	profitability	quality	seasonality	short term reversal	size	value
Intercept	0.003***	0.003***	0.001***	0.001***	0.000**	0.003***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
r_e	0.000	0.003	0.003	0.014**	0.009**	0.001
	(0.000)	(0.003)	(0.003)	(0.006)	(0.005)	(0.001)
Observations	434	434	434	434	434	434
R^2	0.000	0.003	0.003	0.014	0.009	0.001
Adjusted R^2	-0.002	0.001	0.001	0.012	0.007	-0.001
Residual Std. Error	0.000	0.001	0.000	0.002	0.002	0.001
F Statistic	0.001	1.364	1.313	6.046**	3.926**	0.366
Note:						*p<0.1; **p<0.05; ***p<0.01 and (df = 432)

Table 4.7: Regression 2 result. Result of the regression as described in equation (4.10): $\mathbb{E}[f^i_{t+1} \,|\, \mathcal{F}_t] \cdot f^i_{t+1} = \alpha^i_2 + \beta^i_2 \cdot f^i_{t+1} + \epsilon_{t+1}$

4.4 Data description

In this thesis we consider the co-movement of a broad range of macro and financial variables of various frequencies. Macroeconomic data often receive revisions in retrospect by statistical companies, or are potentially released with a delay. We are interested in obtaining a real-time estimate, thus, importantly, the estimates are determined on the real-time data, that were available to investors at the given point in time. Further, to link the monthly model-implied macro variables with the observed quarterly data, we use the time aggregation function of Gomez-Cram (2022) [14].[11] He defines a trend function such that monthly model variables time-aggregate to the observed quarterly growth rates. We further follow Gomez-Cram (2022) [14] in the choice of macroeconomic data. We employ 12 monthly variables and 3 quarterly, which all captures different parts of the economy, these can be seen in Table 4.8. The macro and financial data described here is the data counterpart in Equation 2.1 in Section 2.1 on the business cycle. This data is obtained from the Archival FRED database (AL-FRED) [9], except the 2022 data, which is obtained from Bloomberg Data. The column 'Correlation' indicates the correlation of our data with the data provided in the replication code of Gomez-Cram (2022) [14]. We find that all data series match exactly except for two - why is unknown. Data for the excess market return is obtained from Kenneth R. French Data Library, for a description see [10]. This is the data counterpart of Equation 3.1 in Section 3.1 on excess market return. The sample period for the macroeconomic data runs from January 1950 to September 2022. To account for the extreme macroeconomic values of the Covid-19 pandemic, we introduce an exogenous variable. We argue that this is acceptable, as this model seeks to operate on the financial market and not predict pandemics. We account for the extreme downturns in March/April, as well as the following extreme upturns. Therefore the reader will find data from and including March 2020 up to and including November 2020 to be missing.

[11] Stated in his section on time aggregation on page 941.

Variable name	Ticker	Frequency	Correlation
Industrial Production Index	INDPRO	Monthly	0.79
All Employees, Total nonfarm	PAYEMS	Monthly	> 0.99
Personal Consumption Expendure	PCE	Monthly	> 0.99
Real Personal Income excluding current transfer receipts	W875RX1	Monthly	> 0.99
Indexes of Aggregate weekly hours production and non-supervisory employees, Total Private	AWHI	Monthly	> 0.99
Initial claims	ICSA	Monthly	> 0.99
CAPE		Monthly	0.86
10-Year Treasury Constant Maturity Minus 2-Year Treasury Constant Maturity	T10Y2Y	Monthly	> 0.99
TED spread	TEDRATE	Monthly	> 0.99
Moody's spread	AAA-BAA	Monthly	> 0.99
10-Year Breakeven Inflation Rate	T10YIE	Monthly	> 0.99
VIX		Monthly	> 0.99
GDP		Quarterly	> 0.99
Private Residential Fixed Investment	PRFI	Quarterly	> 0.99
Private Nonresidential Fixed Investment	PNFI	Quarterly	> 0.99

Table 4.8: Data Series Underlying the Business Cycle model

Chapter 5

Conclusion

Excess market returns shows a v-shape in the 6- and 12-month means around recessions, indicating negative excess returns around recession peaks, and strong recovery towards the end of recessions. This evidence suggests that, if we can predict recessions, we should be able to time the market, by selling at the start of recessions and increase risk-taking at a later stage. Using a broad range of macroeconomic growth indicators, we have employed a Gibbs sampling approach, to identify real-time business cycle turning points. To test the timing abilities, we executed a simple timing strategy, which profits by cashing out of the stock market during bad economic times. Here we used the business cycle estimates to indicate a bad economic time. This strategy earns a Sharpe ratio of 82% (52% increase over the buy-and-hold) and a statistically significant annualized CAPM-α of 4.7% (on a 1% significance level).

We extended the analysis to the mean-variance investor, using three one-month-ahead forecasts of excess return: an unconditional historical, a simple OLS forecast and a more sophisticated state-space approach. Our analysis found, most importantly, that the real-time recession probability of our business cycle model caries statistical significant economic relevance (on a 5% significance level). When applying the state-space forecast of excess return, the positive mean-variance CAPM-α became significant (3.37% annualized on a 5% significance level), earned a Sharpe ratio of 64.83% (5% increase over OLS and 18.9% over the unconditional), and a CER difference of 1.56 over the unconditional, indicating that an investor would be willing to pay 156 bps annually to have access to this forecast.

Common for both the simple and dynamic timing strategies were, that they greatly reduced recessionary draw-downs by avoiding the full stock market crashes. However, when examining portfolio return breakdown we found that all timing strategies experienced negative active returns and great draw-downs. We conclude that, even though we produced significant positive alphas, there is room for improvement.

We ended the analysis with a section for future research. We found there to be evidence that the timing ability of our real-time recession probabilities, possibly caries to other factors.

This thesis contributes to the literature by translating the important business cycle algorithm to the modern programming language, Python, as well as providing further evidence on the link between the business cycle and excess return, and the timing abilities of macroeconomic information on the equity market.

Appendix A

Appendix

The code applied in this thesis can be found on the following github link: `https://github.com/cnp777/Thesis-2022.git`.

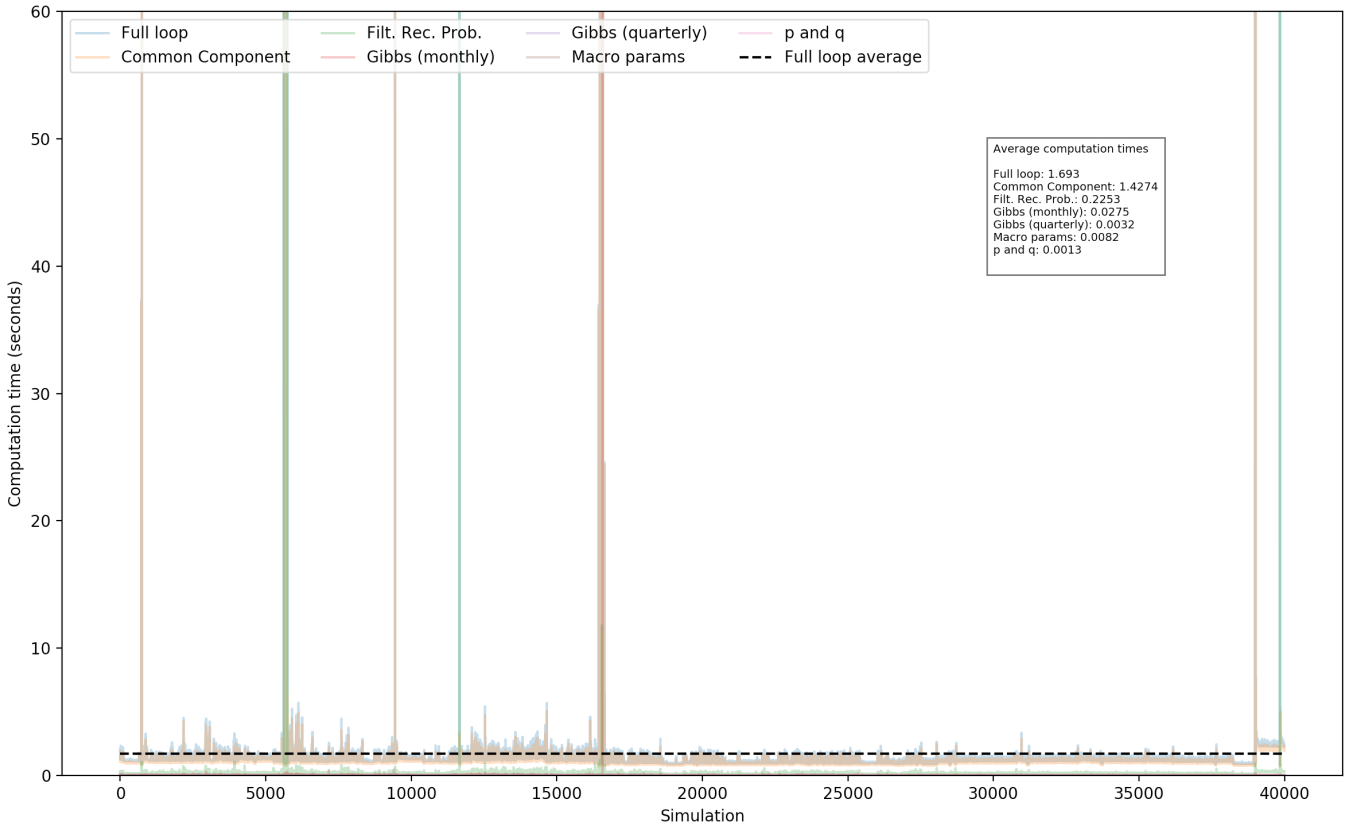


Figure A.1: Computational times *for the functions used in the business cycle algorithm.*

The average computational time for the full loop is approximately 1.5 second. Most computations take well below 1 second, however several computations take well above one minute, pulling the average computation time up.

Bibliography

[1] Andrew Ang. *Asset Management: A Systematic Approach to Factor Investing*. Oxford University Press, 2014. ISBN: 9780199382323.

[2] Andrew Ang, William Goetzmann, and Stephen Schaefer. "Evaluation of Active Management of the Norwegian Government Pension Fund - Global". In: (2009).

[3] John Y. Campbell and John H. Cochrane. "By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior". In: *Journal of Political Economy* 107.2 (1999), pp. 205–251. DOI: 10.1111/jofi.12080.

[4] David Rapach and Guofu Zhou. "Forecasting Stock Returns". In: *Handbook of Economic Forecasting* 2 (2013), pp. 328–383. DOI: 10.1016/B978-0-444-53683-9.00006-2.

[5] J. Durbin and S. J. Koopman. *Time Series Analysis by State Space Methods*. OXFORD STATISTICAL SCIENCE SERIES. Oxford University Press, 2012. ISBN: 978-0-19-964117-8.

[6] Eugene F. Fama and Kenneth R. French. "Business conditions and expected returns on stocks and bonds". In: *Journal of Financial Economics* 25 (1989), pp. 23–49. DOI: https://doi.org/10.1016/0304-405X(89)90095-0.

[7] Eugene F. Fama and Kenneth R. French. "Dividend yields and expected stock returns". In: *Journal of Financial Economics* 22 (1988), pp. 3–25. DOI: https://doi.org/10.1016/0304-405X(88)90020-7.

[8] Vincent Fortuin. "Priors in Bayesian Deep Learning: A Review". In: *International Statistics Review* 90.3 (2022), pp. 563–591. DOI: https://doi.org/10.1111/insr.12502.

[9] Archival FRED. *Economic Data Time Travel from the St. Louis Fed's Economic Research Division*. URL: https://alfred.stlouisfed.org. (accessed: 01.12.2022).

[10] Kenneth French. *Description of Fama/French Factors*. URL: https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/f-f_factors.html. (accessed: 01.12.2022).

[11] Nicolae Gârleanu and Lasse Heje Pedersen. "Dynamic Trading with Predictable Returns and Transaction Costs". In: *The Journal of Finance* LXVIII.6 (2013), pp. 2309–2340. DOI: 10.1111/jofi.12080.

[12] Gary P. Brinson, L. Randolph Hood and Gilbert L. Beebower. "Determinants of Portfolio Performance". In: *Financial Analysts Journal* 42.4 (1986), pp. 39–44. DOI: https://www.jstor.org/stable/4478947.

[13] John Geweke and Gianni Amisano. "Optimal prediction pools". In: *Journal of Econometrics* 164 (2011), pp. 130–141. DOI: https://doi.org/10.1016/j.jeconom.2011.02.017.

[14] Roberto Gómez-Cram. "Late to Recessions: Stocks and the Business Cycle". In: *The Journal of Finance* LXXVII.2 (2022), pp. 923–966. DOI: 10.1111/jofi.13100.

[15] James D. Hamilton. "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle". In: *Econometrica* 57.2 (1989), pp. 357–384. DOI: https://doi.org/10.2307/1912559.

[16] Hanno Lustig and Adrien Verdelhan. "Business cycle variation in the risk-return trade-off". In: *Journal of Monetary Economics* (2012). DOI: http://dx.doi.org/10.1016/j.jmoneco.2012.11.003i.

[17] Nick Heard. *An Introduction to Bayesian Inference, Methods and Computation*. Springer, 2022. ISBN: 978-3-030-82807-3.

[18] Heikki Kauppi and Pentti Saikkonen. "Predicting U.S. Recessions with Dynamic Binary Response Models". In: *The Review of Economics and Statistics* 90.4 (2008), pp. 777–791. URL: https://EconPapers.repec.org/RePEc:tpr:restat:v:90:y:2008:i:4:p:777-791.

[19] Investopedia. *Dotcom bubble*. URL: https://www.investopedia.com/terms/d/dotcom-bubble.asp. (accessed: 01.12.2022).

[20] Investopedia. *Drawdown Definition*. URL: https://www.investopedia.com/terms/d/drawdown.asp. (accessed: 01.12.2022).

[21] Ivo Welch, Amit Goyal. "A Comprehensive Look at The Empirical Performance of Equity Premium Prediction". In: *The Review of Financial Studies* 21.4 (2008), pp. 1455–1508. DOI: https://doi.org/10.1093/rfs/hhm014.

[22] Jiaquan Yao Jian Chen Guohao Tang and Guofu Zhou. "Investor Attention and Stock Returns". In: *SSRN Electronic Journal* (2020). DOI: https://dx.doi.org/10.2139/ssrn.3194387.

[23] John Y. Campbell. "A Variance Decomposition for Stock Returns". In: *The Economic Journal* 101.405 (1991), pp. 157–179. DOI: https://doi.org/10.2307/2233809.

[24] Chang-Jin Kim and Charles R. Nelson. "Business Cycle Turning Points, a New Coincident Index, and Tests of Duration Dependence Based on a Dynamic Factor Model with Regime Switching". In: *The Review of Economics and Statistics* 80.2 (1998), pp. 188–201. DOI: https://www.jstor.org/stable/2646631.

[25] Chang-Jin Kim and Charles R. Nelson. *State-Space Models with Regime Switching*. The MIT Press, 1999. ISBN: 0-262-11238-8.

[26] David A. Levin and Yuval Peres. *Markov Chains and Mixing Times*. URL: https://pages.uoregon.edu/dlevin/MARKOV/. (accessed: 01.12.2022).

[27] Raiden B. Hasegawa Marco Del Negro and Frank Schorfheide. "Dynamic prediction pools: An investigation of financial frictions and forecasting performance". In: *Journal of Econometrics* 192 (2016), pp. 391–405. DOI: https://doi.org/10.1016/j.jeconom.2016.02.006.

[28] Osvaldo Martin. *Bayesian Analysis with Python*. Packt Publishing, 2018. ISBN: 978-1-78934-165-2.

[29] Mehrdad Niknami. "Inflationary Constant Factors and Why Python is Faster Than C++". In: (2019). DOI: 10.48550/arxiv.1911.12338.

[30] Emanuel Moench and Tobias Stein. *Equity premium predictability over the business cycle (Discussion Paper)*. Deutsche Bundesbank, 2021. ISBN: 978-3-95729-833-1.

[31] Alan Moreira and Tyler Muir. "Volatility-Managed Portfolios". In: *The Journal of Finance* LXXII.4 (2017), pp. 1611–1644. DOI: 10.1111/jofi.12513.

[32] Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, and Augusta H. Teller. "Equation of State Calculations by Fast Computing Machines". In: *The Journal of Chemical Physics* 21.6 (1953), pp. 1087–1092. DOI: https://doi.org/10.1063/1.1699114.

[33] Ravin Kumar Osvaldo Martin and Junpeng Lao. *Bayesian Modeling and Computation in Python*. CRC Press, 2022. ISBN: 978-1-003-01916-9.

[34] Thierry Roncalli Pierre Chen Edmond Lezmi and Jiali Xu. "A Note on Portfolio Optimization with Quadratic Transaction Costs". In: *SSRN Electronic Journal* (2019). DOI: 10.2139/ssrn.3683466.

[35] Richard C. Grinold. "The fundamental law of active management". In: *The Journal of Portfolio Management* (1989). DOI: https://doi.org/10.3905/jpm.1989.409211.

[36] Robert A. Jarrow. "Active Portfolio Management and Positive Alphas: Fact or Fantasy?" In: *The Journal of Portfolio Management* 36.4 (2010), pp. 17–22. URL: https://www.proquest.com/scholarly-journals/active-portfolio-management-positive-alphas-fact/docview/742655633/se-2.

[37] Frank Schorfheide. "Learning and monetary policy shifts". In: *Review of Economic Dynamics* Vol. 8 (2005), pp. 392–419. DOI: 10.1016/j.red.2005.01.001.

[38] Towards Data Science. *What does RMSE really mean?* URL: https://towardsdatascience.com/what-does-rmse-really-mean-806b65f2e48e. (accessed: 01.12.2022).

[39] Sean D. Campbell and Francis X. Diebold. "Stock returns and expected business conditions: Half a century of direct evidence". In: *Journal of Business Economic Statistics* 27 (2009), pp. 266–278. DOI: https://www.jstor.org/stable/27799082.

[40] Stackexchange. *Why do we use inverse gamma as prior on variance*. URL: https://stats.stackexchange.com/questions/350924/why-do-we-use-inverse-gamma-as-prior-on-variance-when-empirical-variance-is-gam. (accessed: 01.12.2022).

[41] Masashi Sugiyama. *Introduction to statistical Machine Learning*. Morgan Kaufmann (Elsevir Inc.), 2016. ISBN: 978-0-12-802121-7.

[42] Theis Ingerslev Jensen, Bryan T. Kelly and Lasse Heje Pedersen. "Is There a Replication Crisis in Finance?" In: *Journal of Finance (Forthcoming)* (2022). DOI: https://dx.doi.org/10.2139/ssrn.3774514. URL: https://jkpfactors.com.

[43] Kostas Triantafyllopoulos. *Bayesian Inference of State Space Models. Kalman Filtering and Beyond*. Springer, 2021. ISBN: 978-3-030-76123-3.

[44] Wikipedia. *Beta distribution - Bayesian inference*. URL: https://en.wikipedia.org/wiki/Beta_distribution#Bayesian_inference. (accessed: 01.12.2022).

[45] Wikipedia. *Markov Chain*. URL: https://en.wikipedia.org/wiki/Markov_chain. (accessed: 01.12.2022).

[46] Yves J. Hilpisch. "Computational Finance - Why Python Is Taking Over". In: *Wilmott* 2016.82 (2016), pp. 43–44. DOI: 10.1002/wilm.10489.