Erasmus University Rotterdam

Erasmus University Rotterdam
Erasmus School of Economics
MSc Econometrics & Management Science
Specialisation Quantitative Finance

Master Thesis

Long Term Portfolio Choice

Stock Return Predictability & Regime Uncertainty

Author : Roman Gorlov
Studentnumber : 320372
Supervisor : Karolina Scholtus
Co-reader : Dr. Erik Kole

Contents

Abstract

This Master thesis is on the application of Markov Switching Vector Autoregressive Models (MS-VAR) and their predictability to portfolio choice under regime and parameter uncertainty. The main contribution of this thesis is the extension of such models with Time Varying Transition Probabilities (TVTP). Furthermore their multivariate specifications, their sampling algorithms and marginal ordinal densities are provided. We investigate long term asset allocation for buy-and-hold investors that can allocate to stocks, corporate bonds or cash. We identify two stock market regimes where the first regime has approximately 55% higher annualized volatility relative to the second regime. We use stock market predictors such as Dividend-to-Price Ratio, log Stock Variance and Excess Bond Returns to separate the regimes. By accounting for predictability for stock market regimes we find that these models increase certainty equivalent returns (CER) for buy-and-hold investors with investment horizons starting from 1 year, therefore these type of investors should account for regime predictability in their return forecasting models. This results in higher capital allocations for both stocks and bonds at the expense of cash relative to constant transition probabilities. Buy-and-hold investors with very high risk tolerance may ignore regime predictability if their investment horizon exceeds more than 6 years.

1 Introduction

It is strongly believed by investors, policy makers and academics that the stock market exhibits cyclical behavior of unknown durations. These trends can be upwards or downwards in terms of price changes, commonly known as bull and bear markets. It is not always clear how long these trends persist or how they are formed. We can think of bull and bear markets as different regimes or states of the stock market, as the stock-return distribution parameters change drastically. Practitioners want to know what direction the market is heading and whether one needs to cut his profits and prepare for a bear market. Timely identifying a trend turning point may prevent unnecessary capital losses. Pension funds need to optimize their asset allocation mix for the uncertain long term. Predicting the long term asset returns under uncertainty can help manage the pensions funds' future obligations. Identifying and forecasting these states of the stock market is therefore of great importance for practitioners and academics alike.

In this Master thesis we investigate the effects of long term portfolio choice conditioned on stock market regimes. Latent stock market regimes are identified using a Hidden Markov Model and we account for predictability of stock and bond returns using a VAR model specification. We incorporate predictability for time-varying transition probabilities into the portfolio choice framework under parameter and regime uncertainty. We investigate long term asset allocation for buy-and-hold investors that can allocate to stocks, corporate bonds or cash. We identify two sets of economic regimes for the stock market where the first regime has approximately 55% higher annualized volatility relative to the first regime. Furthermore the predictors exhibit a parameter regime-shift such that the coefficients change substantially. When accounting for regime predictability using log of Stock Variance we find that the conditional variance of cumulative forecasted returns decreases relative to constant transition probabilities. As a result we find that buy-and-hold investors with investment horizon starting from 1 year and upwards should account for regime predictability otherwise they will under-allocate their capital to stocks and bonds for the long term. Their allocation to both stocks and bonds increases when accounting for regime predictability at the expense of cash. Buy-and-hold investors with very high risk tolerance may ignore regime predictability if their investment horizon exceeds more than 6 years.

Stock market regimes are generally assumed to be latent variables, thus they are not truly observed and must be estimated. If we assume regimes to be observable, we can identify them using ex-post dating algorithms, see Pagan and Sossounov [2003] for an example. This method of regime identification relates to finding turning points in the business cycle as show in Bry and Boschan [1971]. More advanced methods that accommodate cyclical time series are the hidden Markov-Switching (MS) models for which transitions between the states are determined by a Markov Chain. Since the states of the stock market are not truly observed we can only obtain probability statements for the latent regime. Hamilton and Lin [1996] used regime-switching models to capture the non-linear dynamics of the stock market and the business cycle. The authors find that this framework proves to be useful both for forecasting stock volatility and for identifying and forecasting economic turning points. Kole and van Dijk [2017] find that parametric methods such a MS model has superior forecasting performance relative to ex-post dating. The study of Guidolin and Timmermann [2007] investigates asset allocation implications in the presence of 4-state regime switching and find that regime switching models capture both short-term and long-term variations in investment opportunities. Ang and Bekaert [2002] report evidence of regimes in stock and bond returns that impact international stock market correlations. They note that regime-switching models are able to capture asymmetric correlations whereas asymmetric GARCH models do not. Maheu, McCurdy, and Song [2012] use Bayesian analysis to estimate a four-state MS model and add restrictions to properly identify bull rallies and bear corrections using weekly index stock returns. Their model strongly dominates non-parametric dating and classical non-restrictive methods in terms of Bayes factors. This motivates the use of MS models to capture cyclicality in the stock market regimes. Maheu and McCurdy [2000] allowed for duration-dependent transition probabilities and duration dependent intra-state dynamics for returns and variance. Their model captures the autoregressive conditional heteroskedasticity structure that many time-series exhibit. They also find that the hazard rate is decreasing with duration for both bull and bear markets indicating a momentum effect. After decoupling conditional returns and conditional variance, it is found that both processes are source of the duration dependence. This motivates that MS models must account for persistence in conditional returns and volatility.

Stock return predictability and the implication for portfolio choice is a major focus for past and present financial research. In the case of independently identically distributed asset returns and constant relative risk aversion, investors are invariant in their investment horizon and their optimal policy is to solve a single period investment problem, see Samuelson et al. [1969]. Studies by Fama and French [1988] and Campbell and Shiller [1988] document strong predictability of stock returns using the dividend-to-price ratio as an economic state variable related to the business cycle. This contrasts the optimal investment policy of single-period optimization and highlights the importance of predictability for long horizon investing. Kandel and Stambaugh [1996] show that asset return predictability is of great importance for the stock demand even on the short-term. After incorporating parameter uncertainty using a Bayesian approach, they noted that stock demand was severely negatively impacted. Campbell and Viceira [2002] show that asset return predictability has great effects on the variance of long-horizon return distribution of stocks and bonds. Barberis [2000] documents the effects of stock return predictability for long term asset allocation. There is enough predictability in stock returns to make investors allocate substantially more to stocks as opposed to t-bills even after accounting for parameter uncertainty, especially for longer investment horizons. Chen [2009] uses various macro-economic variables to forecast the filtered probabilities determined by a simple two-state MS model. He finds that bear markets are more predictable than bull markets, in particular using the yield spread of treasury notes and bills. Macro-economic data seems to contain stronger predictive power for bear markets relative to bull markets. Inflation rates lose their predictive power in the early 1990s as their volatility decreases due to monetary policy changes. Welch and Goyal [2008] challenges the notion of long horizon predictability with an exhaustive study on the empirical performance of equity premium prediction and find that the evidence on predictability is weak and likely to be spurious. Lettau and Van Nieuwerburgh [2008] however note that if we adjust our models to incorporate regime-changes that the stock return predictability significantly improves due to structural shifts in the predictors.

Incorporating parameter uncertainty in portfolio choice theory has been studied by Barberis [2000]. The long horizon effects are studied for investors using a buy-and-hold and dynamic optimal re-balancing strategy with predictability using a VAR model specification. Avramov [2002] investigated model uncertainty by combining return forecasts using various predictor models. But regime uncertainty and especially time-varying transition probabilities are not considered. Guidolin and Timmermann [2005] employ a three-state MS model and find strong implications for the asset allocations using dividend yield as predictor variable within bull and bear market regimes in the UK stock and bond market. They find that investors who ignore predictability in asset returns will face a high utility costs versus incorporating predictability. Guidolin and Timmermann [2006] applied an additional a four-state MS model to fully capture to capture the time-variation in the mean, variance and correlation between large and small firms' stock returns and long-term bond returns. Uniting parameter uncertainty with these above mentioned studies on non-linear asset dynamics allows us to further investigate whether predictability holds for the long horizon investing.

A major restriction for these MS models is that the transition probabilities are assumed to constant. Diebold et al. [1994] extend the original Hamilton [1989] framework by incorporating time-varying transition probabilities that are driven by economic fundamental time series, such as money supplies, relative real outputs and interest rate differentials. Filardo [1994] confirms their methodology by estimating the growth rate in aggregate output with several business-cycle predictors. His findings show great improvement over the fixed-transition probabilities MS model. An empirical issue with forecasting arises due to the data being revised and thus not readily available for real-time forecasters, such that it is not always clear which data to use. Bazzi et al. [2017] propose a dynamic approach to model time variation in transition probabilities for MS models. They let the transition probabilities vary over time as specific transformations of the lagged observations. In particular they use the scores of the predictive likelihood function to model the transition probabilities using the autoregressive score dynamics of Creal et al. [2013]. Such as score uses the information from conditional observations up to time $t-1$ to update the transition probabilities for the following time period t. While this approach is promising, it is unclear whether we can interpret turning points for state-transitions as a result from changes in the time-varying and invariant dynamics of the predictors.

Implementing time-varying transition probabilities in a Bayesian setting introduces another layer of data augmentation and increases the computational burden. A common method for augmentation is that of the random utility model by McFadden et al. [1973]. Holmes et al. [2006] use differences in utilities as latent variables in a Bayesian setting for the binary and Multinomial-Logit model. They sample the Logistic distribution as an infinite scale of mixture normals. As a result, the method is computationally intensive and not preferable for large numbers of iterations and or large models. Frühwirth-Schnatter and Frühwirth [2010] introduce new forms of data augmentation for sampling the parameters of a binary and Multinomial-Logit model from their posterior distribution within a Bayesian setting. The error term of their model is modeled by a finite scale Mixture of Normal distribution. Since this is a finite approximation to the Logistic distribution, it greatly decreases the computational burden without sacrificing accuracy.

A major critique for time-varying transition probability models is that they model the transition probability independent of previously observed states and thus lack regime persistence. As a result, there is no clear interpretation of the variables that drive the switching process. Kaufmann [2015] parametrizes the time-varying transition probability in a centered way that allows the mean-effect of the covariates to influence the time-invariant state persistence. These findings further justify the extension of a time-varying transition probabilities extension in this thesis. We will investigate the implications of predictability of time-varying transition probabilities for long term asset allocation under uncertainty.

The Master thesis is outlined as follows, Section 2 provides an overview of the data, the time period and any data manipulation. Section 3 outlines the Markov Switching Models and extends it with Time-Varying Transition Probabilities. Section 4 documents the Model estimation algorithms under uncertainty and specifies the model priors. Section 5 is on Model comparison and derives the marginal ordinate densities for the sampling algorithms. Section 6 introduces the predictive density and shows the simulation setup for future stock and bond returns. Section 7 presents the Model results, the application for long term asset allocation between stocks and bonds is described in Section 8 and Section 9 concludes.

2 Data

The Welch and Goyal [2008] dataset is used to asses the predictability of economic state variables for excess asset returns in our portfolio choice study under regime and parameter uncertainty. We also use predictors from this dataset to asses predictability for regimes in the portfolio choice study. The economic state variables used can be interpreted as Valuation Ratio's, Bond yield measures and estimates of equity risk. The use of economic state variables is motivated the idea that investors incorporate this information in their decision making as they can be related to the business cycle. We restrict our sample period from July-1952 until and including December-2013. We present the historical time-series of the economic state variables in Figure 1 and their statistical properties are presented in Table 1. The Welch and Goyal [2008] show that is very hard or almost impossible to predict excess asset returns using this dataset. Thus by using this very same dataset and by incorporating the non-linear asset dynamics we hope to show that we can still predict excess asset returns as noted by Lettau and Van Nieuwerburgh [2008] and Guidolin and Timmermann [2006] while at the same maintaining comparability.

Risk-Free Rate The 3-Month Treasury Bill: Secondary Market Rate is assumed to be the risk free rate ($\mathbf{r_{f,t}}$). The data is from the economic research data base at the Federal Reserve Bank at St. Louis (FRED). Since the yield is an annualized figure we use a log-transform to get a monthly rate. Say we have $\mathbf{tbl_t}$, then we construct $\mathbf{r_{f,t}} = \ln(1 + \mathbf{tbl_t}/12)$. Table 1 shows an average risk free rate of 0.38% with a very low standard deviation of 0.25%, it exhibits positive skewness and excess kurtosis due to the high rates in the late 70s and early 80s. This can be seen in Figure 1 where the risk-free rate reached a high of approximately 3%.

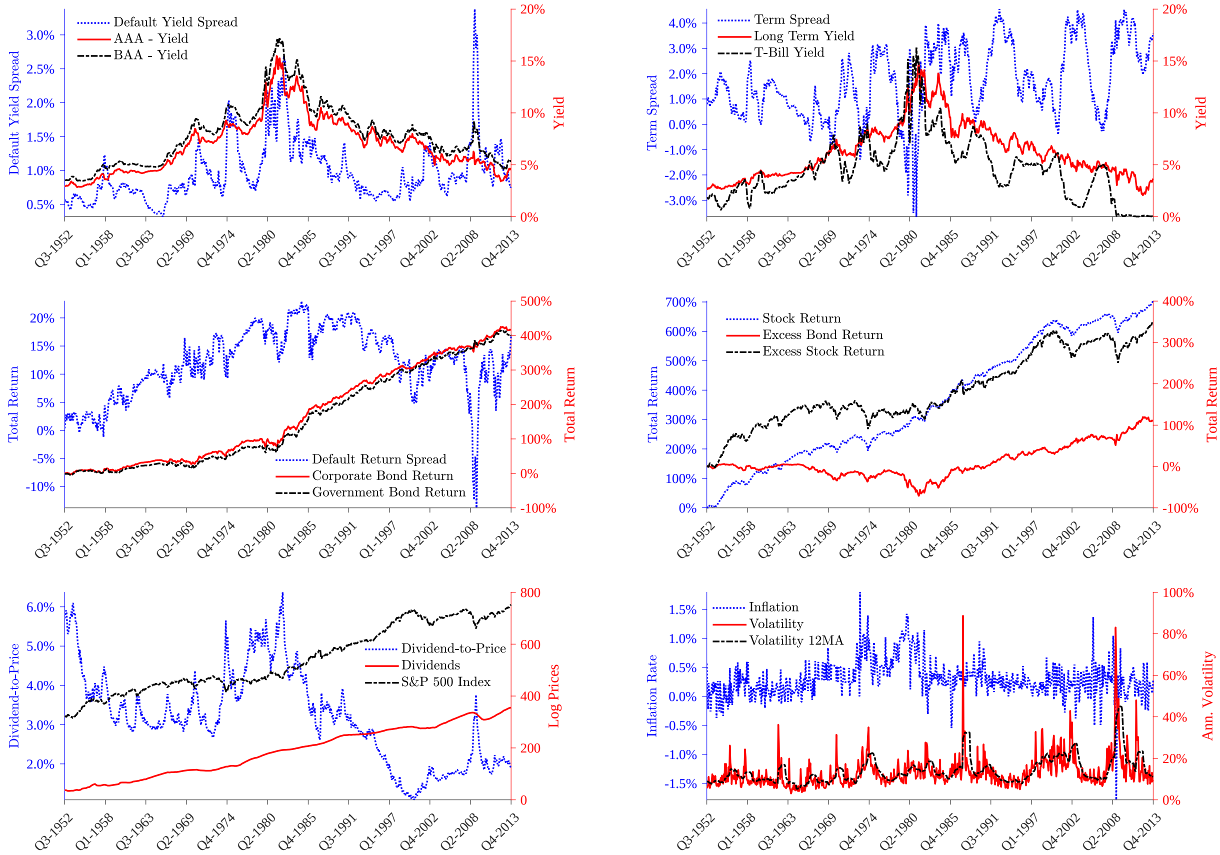

Figure 1: This figure shows historical time-series of the economic state variables from July-1952 until December 2013. The construction of the variables is described in section 2

.

Stock Returns We use the monthly S&P 500 index continuously compounded returns including dividends. The stock returns ($\mathbf{crsp_t}$) are obtained from the Center for Research in Security Press (CRSP) and are month-end values. We compute the log excess stock returns as $\mathbf{r_t} = \ln(1 + \mathbf{crsp_t}) - \mathbf{r_{f,t}}$. Table 1 reports that the average monthly S&P 500 Return was about 0.95% with a standard deviation of 4.21%, which is 14.58% annualized. Stock Returns has a slight negative skew due to several stock market crashes in 1973, 1987, 2000-2002, and 2008 which can be seen in Figure 1.

Stock Variance Stock Variance ($\mathbf{svar_t}$) is computed as the sum of squared daily returns on the S&P 500. We apply regularization in the form of smoothing by applying a 12-month moving average of the squared daily returns. This is computed as $\mathbf{sv_t} = \frac{1}{12}\sum_{k=t-11}^{k=t} \mathbf{svar_k}$. Stock variance has a monthly average of 0.20% which is a higher annualized volatility due to intra-month volatility that cannot be captured with monthly return series. Skewness is strongly positive which indicates many periods of very high risk aversion, which we can see from the 95th percentile which is 3x larger than the mean. Figure 1 depicts stock variance and its smoothed 12-month moving average version over time, we can see some periods of higher stock variances that doesn't necessarily coincide with stock market downturns.

Dividends Dividends are 12-month moving sums of dividends paid on the S&P 500 index. The data is from Robert Shiller's website for the period from January 1927 to December 1987. Data for the period from January 1988 to December 2014 is from the S&P Corporation. The Dividend-to-Price ratio ($\mathbf{dp_t} = ln(\frac{\mathbf{D12_t}}{\mathbf{index_t}})$) is the difference between the log of dividends and the log of the prices for the same time period t. Table 1 depicts the average Dividend-to-Price ratio is about 3.2% per month, which is dividend return of about 0.27% of per month. Figure 1 shows a strongly reducing Dividend-to-Price ratio due to the rise of internet companies with high valuations.

Earnings Earnings are 12-month moving sums of earnings on the S&P 500 index. The data are again from Robert Shiller's website for January 1927 to December 1987. Data for the period from January 1988 to December 2014 is interpolated from quarterly earnings provided by the S&P Corporation. The Earnings-to-Price ratio ($\mathbf{ep_t} = ln(\frac{\mathbf{E12_t}}{\mathbf{index_t}})$) is the difference between the log of earnings and the log of prices. Table 1 shows an average of 6.63% of the Earnings-to-Price and a standard deviation of 2.62% which indicates a strong variability of earnings as can be seen from the 5th and 95th percentiles.

Government Bond Returns We get our Long Term Government Bond Yields ($\mathbf{lty_t}$) from the Ibbotson's Stocks, Bonds, Bills and Inflation Yearbook, which also provides Long Term Rate of Government Bond Returns ($\mathbf{ltr_t}$). The Term Spread ($\mathbf{tms_t} = \mathbf{lty_t} - \mathbf{tbl_t}$) is the difference between the long term yield on government bonds and the Treasury-bill. The Term spread can be seen in Figure 1 exhibiting a strong cyclical pattern with an average of 1.69%. This variable is the only one for which we cannot reject normality. When the term spread widens it represents the steepening of the Yield curve and a flattening vice versa as the term spread tightens. An inverted yield curve is represented by a negative term spread, wherein short term borrowing is more expensive that long term financing, representing a slowdown of the economy.

Corporate Bond Returns Long-term corporate bond returns ($\mathbf{corp_t}$) are again from Ibbotson's Stocks, Bonds, Bills and Inflation Yearbook. Corporate Bond Yields on AAA and BAA-rated bonds are from FRED. The Default Yield Spread ($\mathbf{dfy_t} = \mathbf{BAA_t} - \mathbf{AAA_t}$) is the difference between BAA and AAA-rated corporate bond yields. The Default Return Spread ($\mathbf{dfr_t} = \mathbf{corp_t} - \mathbf{ltr_t}$) is the difference between long-term corporate bond and long-term government bond returns. We depict the total Long-term corporate and government bond return and the total return of the default return spread in Figure 1. The spread is narrow over time, averages 0.02% monthly, but shows a quite some variation (1.40%) over time. The 2008 crash can clearly be seen when the corporate bonds were strongly discounted.

Inflation Inflation ($\mathbf{infl_t}$) is the Consumer Price Index (All Urban Consumers) from the Bureau of Labor Statistics. Because inflation information is released only in the following month, we wait for one month before using it in our monthly regressions.

Table 1: We present the mean, standard deviation, skewness, kurtosis and the HPD estimates of the 5th and 95th percentiles for the variables used in this Master thesis. We restrict our sample period from July-1952 up until December-2013. Furthermore we test each variable for normality using the Jarque-Bera test and provide its statistic. All variables reject normality with the Term Yield Spread as an exception.

	Mean	Std Dev.	Skewness	Kurtosis	5th Perc.	95th Perc.	J-B Statistic
Asset Returns							
Risk Free Rate	0.38%	0.25%	0.863	4.231	0.00%	0.81%	138.098
S&P 500 Return	0.95%	4.21%	-0.406	4.753	-8.32%	8.28%	114.788
Excess Stock Return	0.48%	4.24%	-0.656	5.389	-8.66%	8.09%	228.527
Excess Bond Return	0.15%	2.48%	0.247	7.293	-4.20%	5.65%	574.348
Long Term Government Bond Returns	0.54%	2.79%	0.502	6.232	-5.39%	5.61%	352.212
Long Term Corporate Bond Returns	0.56%	2.49%	0.593	7.871	-3.73%	5.96%	772.790
Default Return Spread	0.02%	1.40%	-0.354	9.960	-2.81%	2.96%	1504.957
Valuation Ratio's							
Dividends	9.871	8.417	0.974	2.981	1.410	27.479	116.752
Earnings	23.285	24.262	1.426	4.027	2.347	82.834	282.542
S&P 500 Index	454.30	499.57	0.986	2.409	23.32	1412.16	130.235
Earnings-to-Price	6.63%	2.62%	0.833	3.501	2.94%	13.33%	93.069
Dividend-to-Price	3.18%	1.19%	0.347	2.524	1.14%	5.41%	21.805
Asset Yield Measures							
AAA - Rated Corporate Bond Yield	6.93%	2.70%	0.747	3.274	2.85%	12.38%	70.990
BAA - Rated Corporate Bond Yield	7.91%	2.98%	0.836	3.526	3.45%	13.95%	94.430
Default Yield Spread	0.98%	0.45%	1.778	7.397	0.34%	1.89%	983.373
Long Term Government Bond Yield	6.30%	2.68%	0.830	3.240	2.48%	12.20%	86.486
US Treasury Bill Yield	4.61%	3.02%	0.872	4.258	0.01%	9.74%	142.156
Term Yield Spread	1.69%	1.43%	-0.118	2.790	-0.46%	4.37%	3.058
Equity Risk & Others							
Stock Return Variance	0.20%	0.41%	10.192	138.279	0.01%	0.62%	575516.575
Stock Return Variance 12MA	0.20%	0.23%	3.900	22.130	0.02%	0.52%	13123.963
CPI Inflation	0.30%	0.31%	0.308	7.189	-0.19%	1.07%	551.206

3 Model Specification

Maheu et al. [2012] provide us with a rich starting ground for further research into regimes and their possible drivers. The presented Markov Switching model, while robust and accurately dating certain market trends, does not allow us to make inferences on the drivers of bull and bear markets. To account for these faults we extend their approach with the Vector Auto Regression (VAR). Such extensions of their model would allow application in long term asset allocation as it gives us insight on the possible drivers of stock returns. Studies done by Guidolin and Timmermann [2005] and Guidolin and Timmermann [2006] suggest that such models can be beneficial in determining long term strategic asset allocation. We will continue to use the Bayesian estimation framework as it allows us to incorporate model and parameter uncertainty in an intuitive way. For further applications of Bayesian Econometrics in Finance please refer to Rachev et al. [2008] and for a full treatise on Markov Switching VAR Models refer to Krolzig [2013].

3.1 Markov Switching Models

We briefly introduce the concept of MS models and the commonly used notation. Then we will introduce the MS model into a VAR framework and thus constitute the MS-VAR model. For a more detailed description we refer the reader to the text by Frühwirth-Schnatter [2006]. Let $R = \{r_1, ..., r_T\}$ be a time series of T-univariate excess stock returns that is generated from a stochastic process r_t. The distribution of r_t depends on the realizations of a hidden discrete stochastic process $S = \{s_1, ..., s_T\}$, which is the regime switching process. The process r_t is directly observable as stock prices are quoted quite frequently, whereas the s_t is a latent random variable, that is only observable indirectly trough its effect on the realizations of r_t. The hidden process s_t is assumed to be an irreducible and aperiodic Markov chain with a discrete state space $\{1, ..., K\}$. This means that we can transition from any state with a single time step to any other state. Its transition process is governed by a square $(K \times K)$ transition probability matrix P. Each element P_{ij} from P is between zero and one, where the i denotes the initial state and j is the target state. The elements of P sum to 1 for each row,

$$P_{ij} = p(s_t = j | s_{t-1} = i), \qquad \sum_{k=1}^{k=K} p_{ik} = 1, \quad \forall i \in K \tag{3.1}$$

3.2 Vector Autoregression Model Framework

We will now extend the univariate MS model from Maheu et al. [2012] to a MS-VAR modeling framework. We initially start with the general VAR(p) model for Excess Stock Returns r_{t+1} and Dividend-to-Price ratio dp_{t+1} that we can easily extend with more variables or more p-lags. The VAR specification allows us model to the joint probability distribution of these processes. It assumes an autoregressive structure such that lagged processes contain predictive information about future realizations of the processes. A single time-period $p = 1$ lag only lags the predictor by 1-time step like in a univariate autoregressive model, here will we focus on a single period lag to keep the amount of model parameters low. The left-hand side of equation (3.2) $y_t = \begin{pmatrix} r_{t+1} \\ dp_{t+1} \end{pmatrix}$ is the column vector containing our variables at time $t + 1$. The right hand side of equation (3.2) consists of the intercept as $c = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$ and the coefficient matrix as $B = \begin{pmatrix} 0 & \beta_1 \\ 0 & \beta_2 \end{pmatrix}$, finally we denote the error term as $\epsilon_{t+1} = (\varepsilon_{1,t+1}, \varepsilon_{2,t+1})$. In result we get the multivariate VAR(1), where again $\epsilon_t \sim i.i.d\ N(0, \Omega)$ with Ω as the error-covariance matrix.

$$y_{t+1} = c + By_t + \epsilon_{t+1} \tag{3.2}$$

Building on top of this framework, let use condition the current VAR(1) parameters on the unobservable Markov process s_t that is governed by transition matrix P from equation (3.1). We condition the full error-covariance matrix and full coefficient matrix B on the unobservable state s_t, such that equation (3.2) becomes,

$$\begin{aligned} y_{t+1} &= c_{s_{t+1}} + B_{s_{t+1}} y_t + \epsilon_{t+1}, \qquad \epsilon_t \sim i.i.d\ N(0, \Omega_{s_{t+1}}) \\ P_{ij} &= p(s_{t+1} = j | s_t = i), \qquad \sum_{k=1}^{k=K} p_{ik} = 1, \quad \forall i \in K \end{aligned} \tag{3.3}$$

For this MS-VAR Model to be properly identified we need to re-order our sampled parameters according some state-identifying restriction. It's also possible to apply random sorting but convergence may take longer, see Maheu et al. [2012], Frühwirth-Schnatter [2006] and Kaufmann [2015] for ordering examples. For our purposes we will explore various stock market states where the stock return variance conditioned on the state parameter differ, such that we order the sampled state vectors by the conditional variance for K-states,

$$Var(r_t | s_t = 1) > Var(r_t | s_t = 2) > ... > Var(r_t | s_t = K) \tag{3.4}$$

In this Master thesis we will explore a 2-state model, with $K = 2$, in order to focus on modeling potential volatility regimes of the aggregate stock market. The excess stock return volatility is assumed to be higher on average in the first regime. We therefore impose the simple following excess stock return volatility restriction, such that $s_t = 1$ will be correctly labeled as a "Risk-On" regime. This allows us to differentiate between market volatility regimes. We do not condition state labeling on duration or some level of expected return. The regime transition matrix is defined as:

$$P = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix} \tag{3.5}$$

We can solve the steady state probabilities of the matrix P using Hamilton [1994]. We impose no restriction on duration but do consider additional restrictions on the conditional excess stock return in the same spirit as Maheu et al. [2012] by using the steady state probabilities derived below,

$$\pi = (A'A)^{-1} A' e \tag{3.6}$$

where $A' = [P' - I, \iota]$ and $e' = [0, 0, 1]$ and $\iota = [1, 1]'$. Using the matrix of the unconditional state probabilities given by equation (3.6), we can compute the long-run excess stock returns in the volatility regimes respectively,

$$E[r_{t+1}|s_{t+1} = 1] = c_1 + \beta_1 dp_t \tag{3.7}$$

$$E[r_{t+1}|s_{t+1} = 2] = c_2 + \beta_2 dp_t \tag{3.8}$$

These computations follow from the fact that the predictors are independent from the current state indicator function and the coefficients are considered here to be a vector of constants. We only use equation (3.4) to identify different volatility regimes.

3.3 Time-Varying Transition Probabilities

As noted earlier, the notion of constant transition probabilities is restrictive and goes against the temporal parameter shifts that we often find in empirical data. Therefore we will also investigate the model extension with Time-Varying Transition Probabilities (TVTP). We follow the methodology of Kaufmann [2015] which is an extension of the difference-in-utility (dRUM) model by Frühwirth-Schnatter and Frühwirth [2010].

In this thesis we will use the logit form to specify time-varying transition probabilities which is a function of the covariate $\tilde{Z}_t$ and a regime specific constant.

$$P(S_t = k|S_{t-1} = l, Z_t, \gamma) = \frac{\exp(Z_t \gamma^z_{lk} + \gamma_{lk})}{\Sigma^K_{j=1} \exp(Z_t \gamma^z_{lj} + \gamma_{lj})} \tag{3.9}$$

The influence of the covariate is decomposed into a time-varying component $Z_t \gamma^z_{lk} = (\tilde{Z}_t - \bar{Z}_t)\gamma^z_{lk}$ which adjusts for the mean effect and a time-invariant average state persistence $\gamma_{lk} = \tilde{\gamma}_{lk} + \bar{Z}\gamma^z_{lk}$. The mean effect thus enters the second component and influences the state persistence. For proper identification of the logit model, we set a reference state $k_0 \in \mathcal{K} = \{1, ..., K\}$ and assume the specific reference state parameters to be zero. Here we set $k_0 = 1$, such that we get transition probability

$$P(S_t = 1|S_{t-1} = l, Z_t, \gamma) = \frac{1}{1 + \Sigma^K_{j=2} \exp(Z_t \gamma^z_{lj} + \gamma_{lj})} \tag{3.10}$$

We can use any variable Z_t to link time-varying transition probabilities to regime switching. However, from a forecasting standpoint it is more preferable to use a predictor that is already included in the MS-VAR model specification.

4 Model Estimation

In this section we discuss the Bayesian estimation prior specification, their respective sampling distributions and the Gibbs sampling algorithm. We follow the methodology of Maheu et al. [2012] for the univariate specification as an example and then introduce the Multivariate specification by leveraging the VAR model of Barberis [2000] to an MS-VAR Model.

4.1 Univariate prior specification

We start by defining our parameter space for the univariate general K state model, $k = 1, ..., K$. There are 3 groups of parameters $M = \{\beta_1, ..., \beta_K\}$, such that $\beta_k = [\beta_{0,k}, \beta_{1,k}]$, $\Sigma = \{\sigma_1^2, ..., \sigma_K^2\}$, and the transition matrix P. Let $\theta = \{M, P, \Sigma\}$ and given data $Y_T = \{y_1, ..., y_T\}$ we augment the parameter space to include the states $S = \{s_1, ..., s_T\}$ so that we sample from the full posterior $p(\theta, S|Y_T)$.

We assume conditionally conjugate priors for each state k, let $\beta_k \sim N(m_k, n_k^2)$, $\sigma_k^{-2} \sim G(v_k/2, w_k/2)$ and each row of P follows a Dirichlet distribution. This specification allows for a Gibbs sampling approach following Chib [1996]. For a standard univariate 2-state model with a single predictor variable and intercept we set flat priors for the prediction equation, but set informative priors for the persistence of regimes in the transition matrix.

$$\beta_k \sim N([0,0], I_2) \tag{4.1}$$

$$\sigma_k^{-2} \sim G(0.5, 0.05) \quad \text{for} \quad k = 1, 2 \tag{4.2}$$

$$\{p_{11}, p_{12}\} \sim Dir(8,2), \{p_{21}, p_{22}\} \sim Dir(2,8) \tag{4.3}$$

These priors should help us uncover the possible prediction equations during empirically relevent regimes.

4.2 Multivariate prior specification

For our multivariate specification a few large changes arise as we have to switch to multivariate distributions to accommodate the VAR specification. We start again by defining our parameter space. There are 3 groups of parameters $M = \{B_1, ..., B_K\}$, $\Sigma = \{\Omega_1, ..., \Omega_K\}$, and the transition matrix P. Let $\theta = \{M, P, \Sigma\}$ and given data $Y_T = \{y_1, ..., y_T\}$ we augment the parameter space to include the states $S = \{s_1, ..., s_T\}$ so that we sample from the full posterior $p(\theta, S|Y_T)$.

We assume conditionally conjugate priors $B_k \sim MN(\mathcal{M}_k, \mathcal{N}_k)$, $\Omega_k^{-1} \sim W(v_k, \mathcal{R}_k)$. For B_k we take a matricvariate normal distribution with mean $\mathcal{M}_k$ and covariance matrix $\mathcal{N}_k$. $\mathcal{M}_k$ is a $(i \times j)$ matrix, for j equations and i regressors that constitute the VAR specification. $\mathcal{N}_k$ is a $(ji \times ji)$ matrix and can also be represented as the covariance matrix $\Omega_k \otimes Q$ where Ω_k is a symmetric $(j \times j)$ covariance matrix and Q is a $(i \times i)$ covariance matrix and is composed of prior data. We will use $\mathbf{I}_j \otimes \mathbf{I}_j$ as starting values for the $\mathcal{N}_k$ covariance matrix.

For Ω_k^{-1} we take a Wishart distribution with degrees of freedom v_k and positive definite scale matrix $\mathcal{R}$. For the VAR 2-state model we set the following flat priors for $\mathcal{N}_k$ to be $\mathbf{O}$ as a $(i \times j)$ matrix of zeros.

$$B_k \sim MN(\mathbf{O}, \mathbf{I}) \ \text{ for } \ k = 1, 2 \tag{4.4}$$

$$\Omega_k^{-1} \sim W(2, \mathcal{R}) \text{ for } k = 1, 2 \text{ and } \mathcal{R} = 0.10 \times \mathbf{I}_j \tag{4.5}$$

4.3 Gibbs sampling algorithm

Gibbs sampling iterates on sampling from the following conditional densities given startup parameters values for M, Σ and P:

$$S|M, \Sigma, P \quad M|\Sigma, P, S \quad \Sigma|M, P, S \quad P|M, \Sigma, S$$

We sample from each of these conditional densities sequentially for each iteration of the Gibbs sampler. Dropping an initial set of draws removes any auto correlation from the startup values, the remaining draw $\{S^{(j)}, M^{(j)}, \Sigma^{(j)}, P^{(j)}\}_{j=1}^{N}$ are collected to estimate features of the posterior density.

Step 1: Sampling the stock market state The first samples step of $S|M, \Sigma, P$ involves a joint draw of all the states. Chib [1996] shows that this can be done by a so-called forward and backward smoother trough the identify

$$p(S|\theta, Y_T) = p(s_T|\theta, Y_T) \prod_{t=1}^{T-1} p(s_t|s_{t+1}, \theta, Y_t) \tag{4.6}$$

Where S is the state vector, θ represents the parameter set and Y_t is the dependent variable. We do not need to draw the states jointly but can in fact draw each state trough a recursion given the previous state. Before we continue it is convenient to adopt the following notation:

$$S_t = \{s_1, ..., s_t\}, \qquad S^{t+1} = \{s_{t+1}, ..., s_n\}$$

and similarly for Y_t and Y^{t+1}. So that, S_t denotes the history up to time t and S^{t+1} the future from $t+1$ to n.

We start by defining the joint conditional distribution which we want to simulate by rewriting it in the following manner using the bayes theorem:

$$\begin{aligned} p(S_n|Y_n, \theta) &= p(s_n|Y_n, \theta) p(S_{T-1}|Y_T, s_T, \theta) \\ &= p(s_n|Y_n, \theta) \times ... \times p(s_t|Y_n, S^{t+1}, \theta) \times ... \times p(s_1|Y_n, S^2, \theta) \end{aligned} \tag{4.7}$$

All terms, except the first, in the above expression have the form of $p(s_t|Y_n, S^{t+1}, \theta)$. Then again by Bayes theorem we rewrite this common form,

$$\begin{aligned} p(s_t|Y_n, S^{t+1}, \theta) &\propto p(s_t|Y_t, \theta) f(Y^{t+1}, S^{t+1}|Y_t, s_t, \theta) \\ &\propto p(s_t|Y_t, \theta) p(s_{t+1}|s_t, \theta) f(Y^{t+1}, S^{t+2}|Y_t, s_t, s_{t+1}, \theta) \\ &\propto p(s_t|Y_t, \theta) p(s_{t+1}|s_t, \theta) \end{aligned} \tag{4.8}$$

The last step follows from the fact that $f(Y^{t+1}, S^{t+2}|Y_t, s_t, s_{t+1}, \theta)$ is independent of s_t. Thus we are left with two terms, one of which is the probability mass function of s_t given (Y_t, θ), and the other is the transition probability of going from s_{t+1} to s_t given θ. The proportionality constant is the sum over s_t, so that we get the recursion or backward smoother:

$$p(s_t|Y_n, S^{t+1}, \theta) = \frac{p(s_t|Y_t, \theta)p(s_{t+1}|s_t, \theta)}{\sum_{s_t} p(s_t|Y_t, \theta)p(s_{t+1}|s_t, \theta)} \tag{4.9}$$

The rest of the problem is solved by determining the conditional probability mass function of the state s at time t. This can be done by a recursion starting with the first observation, this is usually the steady state probability vector from (3.6), so that $p(s_0|Y_0, \theta) = \pi$. The recursion is done in two steps, first with a forward smoother and second with a backward smoother. The forward pass is to compute the Hamilton [1989] filter for $t = 1, ...T$. The filter consists of a prediction step and an update step. Expression (4.10) is called the prediction step, as it predicts s_t on the basis of the information in Y_{t-1} available before time t. The update step (4.11) smooths the probability of s_t conditional on Y_t and θ:

$$p(s_t|Y_{t-1}, \theta) = \sum_k p(s_t|s_{t-1} = k, \theta)p(s_{t-1} = k|Y_{t-1}, \theta) \tag{4.10}$$

$$p(s_t|Y_t, \theta) = \frac{p(s_t|Y_{t-1}, \theta)f(y_t|Y_{t-1}, s_t, \theta)}{\sum_{s_t} p(s_t|Y_{t-1}, \theta)f(y_t|Y_{t-1}, s_t, \theta)} \tag{4.11}$$

Once we have $p(s_T|Y_T, \theta)$ we simulate s_T and can start sampling using the recursion from equation (4.9). With s_T initially sampled, the backward recursion samples states for s_{t-1} using the previously sampled state s_t for $t = T, T-1, ..., 2$.

Step 2 & 3: Sampling the model parameters The second and third sampling steps are more trivial as they use results from the linear regression model. Conditional on S we select the data in regime k and let the number of observations of $s_t = k$ be denoted as T_k, then

$$\begin{gathered} \mu_k|\Sigma, P, S \sim N(a_k, A_k), \\ a_k = A_k \left(\sigma_k^{-2} \sum_{t \in \{t|s_t = k\}} r_t + n_k^{-2} m_k \right), \quad A_k = (\sigma_k^{-2} T_k + n_k^{-2})^{-1} \end{gathered} \tag{4.12}$$

A draw of the variance is taken from

$$\sigma_k^{-2}|M, P, S \sim G\left((T_k + v_k)/2, \left(\sum_{t \in \{t|s_t = k\}} (r_t - \mu_k)^2 + w_k \right)/2 \right) \tag{4.13}$$

Step 4: Sampling the stock market state transition matrix Given the conjugate Dirichlet prior on each row of P, the final step is to sample $P|M, \Sigma, S$ from the Dirichlet distribution. Let each row P_k be the initial state so that $P_k = \begin{pmatrix} p_{k1} & p_{k2} \end{pmatrix}$ with $\sum_{s=1}^{2} p_{ks} = 1$. Then $P_k \sim Dir(\alpha_1, \alpha_2)$ with α_l as the sum of the indicator function whenever a transition is observed from state k to state l plus a prior:

$$\alpha_l = \sum_{t=1}^{T} I(s_t = l | s_{t-1} = k) + \beta_{kl}, \quad l = 1, 2 \tag{4.14}$$

where β_{kl} is the prior for the number of transitions from state k to state l. At each step, each parameter draw is reordered to conform the earlier set restriction from equation (3.4).

4.3.1 Sampling multivariate distributions

The multivariate extension of these sampling steps also follow the linear regression model. These following steps serve to replace steps 3 & 4 in the Gibbs sampling algorithm. Suppose we have predictor variable x_t, which is a $(1 \times k)$ vector including intercept, then we can we store all the data points in the $(T \times k)$ matrix $X = [x_1, \ldots, x_{T-1}]$. We place the response data points in the $(T \times J)$ matrix $Y = [y_2, \ldots, y_T]$. Then conditional on S we select the data in regime k. Such that we need to sample the coefficient matrix B from a matricvariate normal distribution with mean $\widehat{B}_k = (X_k' X_k + Q_k^{-1})^{-1}(X_k' Y_k + Q_k^{-1}\mathcal{M}_k)$ and covariance matrix $\mathcal{V}_k = \Omega_k \otimes (X_k' X_k + Q_k^{-1})^{-1}$ or $\Omega_k \otimes \mathcal{L}_k$.

$$B_k | \Sigma, P, S \sim N(\widehat{B}_k, \mathcal{V}_k) \ \ \forall \ \ t \in \{s_t = k\} \tag{4.15}$$

Which we can, according to van Dijk [2014], estimate by sampling a $(k \times J)$ matrix $\mathbf{U}$, with elements independent and standard normally distributed. Secondly we perform a Choleski decomposition on $\Omega_k = \mathbf{B}'\mathbf{B}$ and $\mathcal{L} = \mathbf{L}'\mathbf{L}$ and compute

$$B_k | \Sigma, P, S = \mathbf{L}'\mathbf{U}\mathbf{B} + \widehat{B}_k \tag{4.16}$$

We sample the inverse error covariance matrix from a Wishart distribution with parameter $\mathcal{R}_k = \left(\mathcal{R}^{-1} + (Y_k - X_k B_k)'(Y_k - X_k B_k)\right)^{-1}$ and $v + T_k - 1$ degrees of freedom.

$$\Omega_k^{-1} | M, P, S \sim W(\mathcal{R}_k, v + T_k - 1) \tag{4.17}$$

Which we can again sample by sampling $(v + T_k - 1 \times J)$ matrix $\mathbf{U}$, with elements independent and standard normally distributed. Next we perform Choleski decomposition on $\mathcal{R}_k = \mathbf{A}'\mathbf{A}$ and compute

$$\Omega_k^{-1} | M, P, S = \mathbf{A}'(\mathbf{U}'\mathbf{U})\mathbf{A} \tag{4.18}$$

4.3.2 Sampling Time-Varying Transition Probabilities

In this section we describe the sampling scheme for the logit model specification of the time-varying transition probabilities. The derivations follow Kaufmann [2015] which are largely based on Frühwirth-Schnatter and Frühwirth [2010]. The distribution of interest is that of $p(\gamma | S, Z)$, which is the conditional posterior distribution of γ conditioned on joint draw of all states and the predictor variable Z. We start by introducing the random utility model as the representation for the logit model.

$$\begin{aligned} S^u_{k,t} &= Z_t'\gamma_k + \nu_{k,t}, \quad \forall k \in \mathcal{K}_{-k_0} \\ S^u_{k_0,t} &= \nu_{k_0,t}, \quad \text{since } \gamma_{k_0} = 0 \end{aligned} \tag{4.19}$$

Here we model the regime specific utility as a function of the predictor variable Z_t with γ_k the estimator of interest. We have state k_0 as our reference state. We can rewrite this model as the difference in utility against our reference state, i.e. the dRUM model.

$$\begin{aligned} S^u_{k,t} - S^u_{k_0,t} &= Z_t'\gamma_k + \nu_{k,t} - \nu_{k_0,t} \\ s_{k,t} &= Z_t'\gamma_k + \epsilon_{k,t}, \quad \epsilon_{k,t} \sim Logistic, \;\; \forall k \in \mathcal{K}_{-k_0} \end{aligned} \tag{4.20}$$

The error terms in equation (4.20) are not independent any more across states as they follow a multivariate logistic distribution. This complicates the MCMC sampling method unfavourably. We consider the partial representation of the dRUM model described by Frühwirth-Schnatter and Frühwirth [2010]. This relies on the observation that

$$S_t = k \Leftrightarrow S^u_{k,t} > S^u_{-k,t}, \qquad S^u_{-k,t} = \max_{j \in \mathcal{K}_{-k}} S^u_{j,t} \tag{4.21}$$

Here we observe that the state at time t is k which is equivalent to the fact that the utility for state k at time t is greater than all other utilities for states other than k. Such that the maximum utility is observed for the observed state, which makes sense considering the Hamilton filter. We must note that the reference state k_0 is a candidate for the maximum. Next we define the difference of the observed state utility with all other utilities but the reference state as $\omega_{k,t}$. Here we consider D^k_t as the indicator variable for our state observations.

$$\omega_{k,t} = S^u_{k,t} - S^u_{-k,t}, \quad D^k_t = I(S_t = k), \quad \forall k \in \mathcal{K}_{-k_0} \tag{4.22}$$

The difference in utility $\omega_{k,t}$ has an explicit form within the multinomial logit model. As the maximum utility is obtained for the observed state k, such that $\exp(-S^u_{-k,t})$ is the minimum value among other possible states $\exp(-S^u_{j,t})$, which are individually exponentially distributed with parameter $\lambda_{j,t} = \exp(Z_t'\gamma_j)$. The minimum then follows an exponential distribution with the sum of the individual parameters.

$$\begin{aligned} \exp(-S^u_{j,t}) &\sim \mathcal{E}(\lambda_{j,t}) \\ \min_{j \in \mathcal{K}_{-k}} \exp(-S^u_{j,t}) &\sim \mathcal{E}\left(\sum_{j \in \mathcal{K}_{-k}} \lambda_{j,t}\right) \\ \exp(-S^u_{-k,t}) &\sim \mathcal{E}\left(\lambda_{-k,t}\right) \end{aligned} \tag{4.23}$$

where $\lambda_{-k,t} = \sum_{j \in \mathcal{K}_{-k}} \lambda_{j,t}$. We may rewrite $S^u_{-k,t} = \log(\lambda_{-k,t}) + \nu_{-k,t}$ where $\nu_{-k,t}$ follows an Extreme Value distribution. The multinomial logit model now has the partial dRum representation

$$\begin{aligned}\omega_{k,t} &= S^u_{k,t} - S^u_{-k,t} = Z_t'\gamma_k - log(\lambda_{-k,t}) + \nu_{k,t} - \nu_{-k,t} \\ &= Z_t'\gamma_k - \log(\lambda_{-k,t}) + \epsilon_{k,t}, \qquad \epsilon_{k,t} \sim i.i.d\ Logistic\end{aligned} \tag{4.24}$$

The constant $-\log(\lambda_{-k,t})$ depends only on the parameters γ_{-k}, thus given $\omega_k^T = \{\omega_{k,1}, \ldots, \omega_{k,T}\}$ and γ_k we obtain a linear regression with parameter γ_k and logistic errors.

We can now describe the sub-sampling steps more closely. For each state k we first sample the latent utility differences ω_k^T from the truncated logistic distributions truncated to $[0, \infty)$ if $S_t = k$ and truncated to $(-\infty, 0]$ if $S_t \neq k$. We draw T values $W_{k,t}$ from a uniform distribution $W_{k,t} \sim U(0,1)$ and set

$$\omega_{k,t} = Z_t'\gamma_k - \log(\lambda_{k,t}) + F_\epsilon^{-1}\left(D_t^k + W_{k,t}\left(1 - D_t^k - \pi_{k,t}\right)\right). \tag{4.25}$$

where $\pi_{k,t} = P(D_t^k = 1|\gamma) = 1 - F_\epsilon(-Z_t'\gamma_k + \log(\lambda_{-k,t}))$. F_ϵ is the CDF of the logistic distribution which we compute using the hyperbolic tangent, such that $F_\epsilon(x;\mu,s) = \frac{1}{2} + \frac{1}{2}\tanh(\frac{x-\mu}{2s})$. F_ϵ^{-1} is the inverse of the logistic CDF such that $F_\epsilon^{-1}(p) = \log(p) - \log(1-p)$.

Given $\omega^{K,T}$, the posterior of γ_k is derived based on equation (4.24). We need to approximate the logistic distribution of the errors $\epsilon_{k,t}$ by a mixture of normal distributions with M components. The components are denoted by $R_{k,t}$ and are drawn from a multinomial distribution.

$$P(R_{k,t} = r|\omega_{k,t}, \gamma_k) \propto \frac{w_r}{s_r}\exp\left\{-\frac{1}{2}\left(\frac{\omega_{k,t} + \log(\lambda_{-k,t}) - Z_t'\gamma_k}{s_r}\right)^2\right\} \tag{4.26}$$

where $r = 1, \ldots, 6$, and the respective component's standard deviation s_r and weight w_r, are taken from Frühwirth-Schnatter and Frühwirth [2010] Table 1. The mixture of these components approximates the logistic distribution efficiently and accurately. Having sampled the components, our model in equation (4.24)becomes normal in γ_k.

$$\tilde{\omega}_{k,t} = \omega_{k,t} + \log(\lambda_{-k,t}) = Z_t'\gamma_k + \epsilon_{k,t}, \qquad \epsilon_{k,t}|R_{k,t} \sim N(0, s^2_{R_{k,t}}) \tag{4.27}$$

Assuming a normal prior for $\gamma_k, p(\gamma_k) = N(g_0, G_0)$, and conditional on ω_k^T and R_k^T the posterior is also normal

$$\begin{gathered}p(\gamma_k|\omega_k^T, R_k^T) = N(g_k, G_k) \\ g_k = G_k\left(\sum_{t=1}^T Z_t\tilde{\omega}_{k,t}/s^2_{k,t} + G_0^{-1}g_0\right) \qquad G_k = \left(\sum_{t=1}^T Z_tZ_t'/s^2_{k,t} + G_0^{-1}\right)^{-1}\end{gathered} \tag{4.28}$$

4.4 Posterior estimates

Simulation consistent estimates can be obtained as sample averages of the draws. For example, the posterior mean of the state dependent mean, covariance matrix of returns and transition probability matrix are estimated as

$$\frac{1}{N}\sum_{j=1}^{N} M_k^j, \quad \frac{1}{N}\sum_{j=1}^{N} \Sigma_k^j, \quad \frac{1}{N}\sum_{j=1}^{N} P^j \tag{4.29}$$

for $k = 1, ..., K$ and simulation consistent estimates of $E[\mu_k|Y_T]$ and $E[\sigma_k|Y_T]$ respectively. Another important byproduct of Gibbs sampling is an estimate of the smoothed state probabilities $p(s_t|Y_T)$ which can be used for investor inference, it is estimated as the average across total draws when state k is drawn:

$$p(s_t \widehat{= k}|Y_T) = \frac{1}{N}\sum_{j=1}^{N} I(s_t = k, S^j) \tag{4.30}$$

Next to the parameter estimates we need to be able to make some useful inferences about the persistence of stock market regimes. We can compute the expected duration using the transition probability matrix P. Let the duration D_k be the number of time periods in the k^{th} regime. Then using the Markov property we can compute the probability of the duration being h time periods as

$$P(D_k = h) = p_{kk}^{h-1}(1 - p_{kk}) \tag{4.31}$$

Then the expected duration of regime k will be the sum of the geometric series

$$E[D_k] = \sum_{h=0}^{\infty} hP(D_k = h) = \frac{1}{1 - p_{kk}} \tag{4.32}$$

Another useful inference we can draw is the expected cumulative return of regimes, this is simply the expected duration for regime k multiplied with its respective expected return.

$$\bar{r}_k = \mu_k E(D_k) \tag{4.33}$$

We can decompose the conditional expected variance per regime into a mean and variance part

$$Var(r_t|s_t) = Var(E(r_t|s_t)|s_t) + E(Var(r_t|s_t)|s_t) \tag{4.34}$$

By computing the ratio between the variance of the expectation with the variance we get a measure of explained variance

$$\frac{Var(E(r_t|s_t)|s_t)}{Var(r_t|s_t)} \tag{4.35}$$

5 Model Comparison

5.1 Bayes factors

We can compare models using Bayes factors, however we need to be able to compute the marginal likelihoods for this to be feasible. Bayes factors allows us to compare non-nested models as well as specifications with a different number of states. The advantage of Bayes factors is that it penalizes model that tend to over-fit while not providing improved predictions. The general Markov-switching model with K states has their marginal likelihood for model M_i defined as follows:

$$p(r|M_i) = \int p(r|M_i, \theta)p(\theta|M_i)d\theta \tag{5.1}$$

Equation (5.1) integrates out parameter uncertainty with $p(\theta|M_i)$ as the prior and equation (5.2) is the likelihood which has S integrated out of the return probability distribution function using equation (5.3)

$$p(r|M_i, \theta) = \prod_{t+1}^{T} f(r_t|Y_{t-1}, \theta) \tag{5.2}$$

$$f(r_t|Y_{t-1}, \theta) = \sum_{k=1}^{K} f(r_t|Y_{t-1}, \theta, s_t = k)p(s_t = k|\theta, Y_{t-1}) \tag{5.3}$$

We obtain the term $p(s_t = k|\theta, Y_{t-1})$ from the Hamilton filter mentioned in equation (4.10). We can estimate (5.1) according to Newton and Raftery [1994] as :

$$p(r|M_i) = \left\{ \frac{1}{G} \sum_{g=1}^{G} \left(\frac{1}{f(r|\theta_i^{(g)}, M_i)} \right) \right\}^{-1} \tag{5.4}$$

which is the harmonic mean of the likelihood values. Chib [1995] mentions that although this estimate is a simulation-consistent estimate of $p(r|M_i)$, it is not stable, because the inverse likelihood does not have a finite variance.Chib [1995] further shows how to estimate the marginal likelihood for Markov-switching models. His estimate is based on re-arranging Bayes' theorem as

$$p(\widehat{r|M_i}) = \frac{p(r|M_i, \theta^*)p(\theta^*|M_i)}{p(\theta^*|r, M_i)} \tag{5.5}$$

Where θ^* is a point of high mass in the posterior probability distribution function. the terms in the numerator are directly available above while the denominator can be estimated using additional Gibbs sampling runs.

5.2 Marginal ordinate densities

We follow Chib [1995] and leave out the model parameter M_i for ease of notation, then since $\theta = \{M, \Sigma, P\}$ we get using Bayes' theorem:

$$\begin{aligned} p(\theta^*|r) &= p(M^*, \Sigma^*, P^*|r) \\ &= p(M^*|r)p(\Sigma^*|M^*, r)p(P^*|M^*, \Sigma^*, r) \end{aligned} \tag{5.6}$$

We rewrite each of the product terms on the right hand site in (5.6) separately so that we can determine the approximation of these conditional densities.

$$\begin{aligned} p(M^*|r) &= \int p(M^*, \Sigma, P, S|r) d\Sigma dP dS \\ &= \int p(M^*|\Sigma, P, S, r) p(\Sigma, P, S|r) d\Sigma dP dS \\ &\approx \frac{1}{G} \sum_{g=1}^{G} p(M^*|\Sigma^{(g)}, P^{(g)}, S^{(g)}, r) \end{aligned} \tag{5.7}$$

The term (5.7) is the marginal ordinate density, which can be estimated from the draws of the initial Gibbs run.

$$\begin{aligned} p(\Sigma^*|M^*, r) &= \int p(\Sigma^*, S, P|M^*, r) dP dS \\ &= \int p(\Sigma^*|S, P, M^*, r) p(S, P|M^*, r) dS dP \\ &\approx \frac{1}{G} \sum_{g=1}^{G} p(\Sigma^*|S^{(g)}, P^{(g)}, M^*, r) \end{aligned} \tag{5.8}$$

The term (5.8) is the reduced conditional ordinate density which we can estimate by sampling draws from the so called reduced Gibbs sampler. First we sample from the reduced complete conditional density $p(S|M^*, \Sigma, P)$ while setting $M = M^*$. Here M^* is our posterior mean from the initial full Gibbs sampler. Next we sample from the complete conditional density $p(P|M^*, \Sigma, S)$ and again set $M = M^*$ and finally we sample $p(\Sigma|M^*, P, S)$ and set $M = M^*$. Chib [1995] mentions that we need less sampling runs because we have one less parameter to worry about so the sampler converges faster for the $p(\Sigma^*|M^*, r)$ approximation.

$$\begin{aligned} p(P^*|M^*, \Sigma^*, r) &= \int p(P^*, S|M^*, \Sigma^*, r) dS \\ &= \int p(P^*|S, M^*, \Sigma^*, r) p(S|M^*, \Sigma^*, r) dS \\ &\approx \frac{1}{G} \sum_{g=1}^{G} p(P^*|S^{(g)}, M^*, \Sigma^*, r) \end{aligned} \tag{5.9}$$

For the third and last term in (5.6) we sample more draws from a second reduced Gibbs sampler. Here we start with sampling from the complete conditional density $p(S|M^*, \Sigma^*, P)$ while setting $M = M^*$ and $\Sigma = \Sigma^*$. Next we sample from the complete conditional density $p(P|M^*, \Sigma^*, S)$ and again set $M = M^*$ and $\Sigma = \Sigma^*$. Collecting terms (5.7), (5.8), (5.9) and the numerator terms in (5.5) we can compute the marginal likelihood as follows

$$\begin{aligned}\ln \widehat{p(r|M_i)} = &\ln p(r|M_i, \theta^*) + \ln p(\theta^*|M_i)\\ &- \ln \frac{1}{G}\sum_{g=1}^{G} p(M^*|\Sigma^{(g)}, P^{(g)}, S^{(g)}, r)\\ &- \ln \frac{1}{G}\sum_{g=1}^{G} p(\Sigma^*|S^{(g)}, P^{(g)}, M^*, r)\\ &- \ln \frac{1}{G}\sum_{g=1}^{G} p(P^*|S^{(g)}, M^*, \Sigma^*, r)\end{aligned} \tag{5.10}$$

then the log-Bayes factor between model M_i and M_j is defined as

$$\ln\ (BF_{ij}) = \ln\ (p(r|M_i)) - \ln\ (p(r|M_j)) \tag{5.11}$$

We use the Kass and Raftery (1995) interpretation of evidence for M_i versus M_j as:

$$\begin{aligned} 0 \leq \mathbf{ln}\ (BF_{ij}) < 1 &: \text{ Not worth more than a bare mention.}\\ 1 \leq \mathbf{ln}\ (BF_{ij}) < 3 &: \text{ Positive evidence.}\\ 3 \leq \mathbf{ln}\ (BF_{ij}) < 5 &: \text{ Strong evidence.}\\ \mathbf{ln}\ (BF_{ij}) \geq 5 &: \text{ Very strong evidence.}\end{aligned}$$

6 Predictive Density

Crucial to theoretical and empirical validation of our Bayesian - MS VAR model is that a predictive density of future returns can be computed that integrates out all current information and parameters. The predictive density for expected excess returns based on all current information at time t is computed as

$$p(y_{t+1}|Y_t) = \int f(y_{t+1}|Y_t, \theta)p(\theta|Y_t)d\theta \tag{6.1}$$

We can decompose the conditional density of future returns based on current information into independent components conditional on the future stock market state.

$$f(y_{t+1}|Y_t, \theta) = \sum_{k=1}^{K} f(y_{t+1}|Y_t, \theta, s_{t+1} = k)p(s_{t+1} = k|\theta, Y_t) \tag{6.2}$$

It is trivial to see that the probability of the future stock market state is another form of the prediction step from the Hamilton filter of equation (4.10). Following this result we can use the Gibbs sampling draws $\left\{S^j, M^j, \Sigma^j, P^j\right\}_{j=1}^{N}$ based on data Y_t to approximate the predictive density as

$$p(\widehat{y_{t+1}|}Y_t) = \frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K} f(y_{t+1}|Y_t, \theta^{(i)}, s_{t+1} = k)p(s_{t+1}^{(i)} = k|\theta^{(i)}, Y_t) \tag{6.3}$$

where $f(y_{t+1}|Y_t, \theta^{(i)}, s_{t+1} = k)$ is the draw from $N(\widehat{B}_k^{(i)}, \mathcal{V}_k^{(i)})$ and $p(s_{t+1}^{(i)} = k|\theta^{(i)}, Y_t)$ is the transition probability.

6.1 Simulating from the Markov switching model

For inferences on long time periods out-of-sample we want to be able to to simulate from the Bayesian MS-VAR model. The simulating paradigm will then consist of simulating J paths with each length τ. Suppose we are in path j and we sample state k at time period T. We assume that we do not know the parameters that make up every path.

Step 0: draw $\Omega_k^{-1,(j)} \sim W(\mathcal{R}_k, v + T_k - 1)$ and $B_k^{(j)} \sim N(\widehat{B}_k^{(j)}, \Omega_k^{(j)})$ for $k = 1, 2, 3, 4$.

Step 1: draw $S_{T+1}^{(j)}$ from the distribution $p(s_{T+1} = k|s_T, \theta^{(i)}, Y_T)$ and ε_{T+1} from $N(\mathbf{0}, \Omega_k^{(j)})$. Update the information and set

$$y_{T+1}^{(j)} = c_k + B_k^{(j)} y_T + \epsilon_{k,T+1}^{(j)} \tag{6.4}$$

Step τ: draw $S_{T+\tau}^{(j)}$ from the distribution $p(s_{T+\tau} = k|s_{T+\tau-1}, \theta^{(i)}, Y_{T+\tau-1})$ and $\varepsilon_{T+\tau}$ from $N(\mathbf{0}, \Omega_k^{(j)})$. Update the information and set

$$y_{T+\tau}^{(j)} = c_k + B_k^{(j)} y_{T+\tau-1} + \epsilon_{k,T+\tau}^{(j)} \tag{6.5}$$

Thus we draw the model parameters only once for each path, however, we draw the future state conditional on the previously drawn information for each time period.

6.2 Simulating from the Markov Switching VAR-TVTP Model

Simulation from the Markov Switching VAR-TVTP model is not much different than from the previous section. Again we simulate J paths with each length τ, but the simulation of the sampled state is now governed by the difference in utility model (dRUM). We recall from equation (4.28) that the dRUM model has a normally distributed posterior, such that we can simply use the drawn γ_k from the Gibbs sampler. However at the end of each step we construct the transition probability matrix ξ, such that we have a time-varying transition matrix in each path j.

Step 0: draw the coefficient covariance matrix $\Omega_k^{-1,(j)} \sim W(\mathcal{R}_k, v + T_k - 1)$, the coefficient matrix $B_k^{(j)} \sim N(\widehat{B}_k^{(j)}, \Omega_k^{(j)})$ and the regime covariates $\gamma_k^{(j)} \sim N(g_k^{(j)}, G_k^{(j)})$ for all k states. Use the regime covariates to construct the transition matrix ξ_T.

Step 1: draw a random state $S_{T+1}^{(j)}$ from the transition matrix $\xi_T|s_T, \theta^{(i)}, Y_T$ and given the drawn state k draw the error covariance matrix ε_{T+1} from $N(\mathbf{0}, \Omega_k^{(j)})$. Update the information set, compute the

VAR equation and transition matrix ξ_{T+1}:

$$y_{T+1}^{(j)} = c_k + B_k^{(j)} y_T + \epsilon_{k,T+1}^{(j)} \tag{6.6}$$

Step τ: draw a random state $S_{T+\tau}^{(j)}$ from transition matrix $\xi_{T+\tau-1}|s_{T+\tau-1}, \theta^{(i)}, Y_{T+\tau-1}$ and given the drawn state k draw the error covariance matrix $\varepsilon_{T+\tau}$ from $N(\mathbf{0}, \Omega_k^{(j)})$. Update the information set and compute the VAR equation and transition matrix $\xi_{T+\tau}$:

$$y_{T+\tau}^{(j)} = c_k + B_k^{(j)} y_{T+\tau-1} + \epsilon_{k,T+\tau}^{(j)} \tag{6.7}$$

7 Portfolio Choice

We follow the methodology of optimal portfolio choice from Barberis [2000] and Brandt [2010]. The goal is to determine a fixed asset allocation that will be optimal for some fixed investment horizon. We consider a buy-and-hold investor with an investment horizon of τ months. The buy-and-hold investor can invest in two continuously compounded risky assets and a risk-free asset at the beginning of his investment period. The risk-free rate R_f asset is assumed to have a constant zero return thus constituting a cash asset. He maximizes his conditional expectation of utility, which is a function of his terminal wealth $W_{t+\tau}$, with respect to the portfolio weights at beginning of his investment period $\omega = \{\omega_s, \omega_b, \omega_{r_f}\}$ and all information I_t known up to time t. If initial wealth $W_t = 1$ and ω is the allocation to stocks w_s, bonds w_b and the risk-free asset w_{rf}, then his terminal wealth is given by

$$W_{t+\tau} = \omega_{r_f} \exp(r_f \tau) + \omega_s \exp(r_{t->t+\tau}) + \omega_b \exp(corp_{t->t+\tau}) \tag{7.1}$$

$r_{t->t+\tau}$ denotes the cumulative stock return from to beginning of the investment period up to investment horizon τ, the same holds for the cumulative bond return $corp_{t->t+\tau}$. We ignore taxes on capital gains, transaction costs or slippage costs and also rule out short selling. We further assume that the investor has constant relative risk aversion (CRRA) power utility with risk aversion γ, such that he considers wealth losses to be a greater utility loss as opposed to wealth gains. The utility function with terminal wealth is then defined as

$$U(W_{t+\tau}) = \frac{W_{t+\tau}^{1-\gamma}}{1-\gamma} \tag{7.2}$$

The buy-and-hold investor must then maximize his conditional expectation of utility based on his terminal wealth

$$\max_{\omega} E_t \left[\frac{\{\omega_{r_f} \exp(r_f \tau) + \omega_s \exp(r_{t->t+\tau}) + \omega_b \exp(corp_{t->t+\tau})\}^{1-\gamma}}{1-\gamma} \right] \tag{7.3}$$

The buy-and-hold investor computes his conditional expected utility at the beginning of his investment period t as denoted by E_t. The predictive density for the τ-period cumulative asset returns, $r_{t->t+\tau}$, can be constructed using the iterative scheme from section 6. To recall, the predictive density for the joint distribution of stocks and bonds $z_{t->t+\tau} = (r_{t->t+\tau}, corp_{t->t+\tau})$ is $p(z_{t->t+\tau}|Y_t)$, where Y_t is the data observed by the investor up until the start of his investment period. We must also account for estimation risk in this optimization problem by incorporating the posterior distribution of the parameters θ, so the optimization problem then becomes

$$\max_{\omega} \int U(W_{t+\tau}) p(z_{t->t+\tau}|Y_t, \theta) p(\theta|Y_t) dz_{t->\tau} d\theta \tag{7.4}$$

We can approximate this integral by

$$\frac{1}{J} \sum_{j=1}^{J} \frac{1}{1-\gamma} \left[\omega_{r_f} \exp(r_f \tau) + \omega_s \exp(r^j_{t->t+\tau}) + \omega_b \exp(corp^j_{t->t+\tau}) \right]^{1-\gamma} \tag{7.5}$$

To solve this buy-and-hold portfolio we need to sample $J \times \tau$ asset return-paths using our two-state MS model and choose ω that maximizes the investor's utility for varying investment horizons. Each path will consist of expected excess asset returns τ forecasts conditional on the state draws. This problem is solved by using the simulated excess asset returns from each iteration of the Gibbs sampler.

8 Model Results

In this section the results of both models are presented and evaluated. First the Markov-Switching Vector Auto Regressive Model is discussed, or more concise MS-VAR Model. Secondly the Markov-Switching with Time Varying Transition Probabilities Vector Auto Regressive Model is evaluated, denoted as the MS-VAR-TVTP Model. The posterior distributions of both models' parameters are sampled and evaluated, specifically the posterior distribution model covariance and regime characteristics. The posterior distribution of the transition and state probabilities are also investigated. We regress the Excess Stock Returns, Stock Variance, Dividend-to-Price Ratio and Excess Bond Returns on an intercept, lagged Stock Variance, lagged Dividend-to-Price Ratio and finally lagged Excess Bond Returns. The variables are lagged by a single period, i.e. one month (one data-point) in the time-series. Other variables such Yield Default Spread, Default Return Spread or the Term Spread that can proxy some latent stock market risk factor can be used seperate stock market regimes(result not shown) in the same manner as Stock Variance, but the use of Stock Variance is more intuitive.

The posterior distribution is estimated from 30,000 draws and has a burn-in of the first 10,000 draws, the thin value is set to 10. State draws were permuted on the estimated Variance of Excess Stock Returns, from high to low. Such that economic state $S_t = 1$ can be considered a "Risk-On" state, representing a period of high stock return variance. In such periods the uncertainty of future excess stock returns increases and thus decreasing the risk appetite of investors. The tables contain means, medians, standard deviations and the highest probability densities (HPD) of the posterior distributions, see Chen and Shao [1999] for further details. The HPD is considered to be the equivalent of credibility intervals of estimated model parameters in a frequentist framework.

8.1 Markov Switching VAR Model

In Table 2 the posterior distribution of the model parameters are presented for both states. In Table 3 the posterior distribution of the covariance matrix is presented for both regimes. Both tables are accompanied by Figure 2 for the posterior distribution of the model parameters and Figure 3 for the posterior distribution of the covariance matrix. The regime characteristics are shown in Table 4 and is accompanied by Figure 4. We first discuss the posterior distribution of the model coefficients, then the posterior distribution of the covariance matrix and finally the posterior distribution of the regime characteristics.

We see a strong shift in the posterior distribution of Excess Stock Return predictors in Table 2. The intercept of Excess Stock Returns r_{t+1} nearly halves from 8.897 to 4.005 when switching from regime $s_t = 1$ to regime $s_t = 2$, concurrently the std. dev also more than halves when switching regimes. This indicates a more than double the unconditional excess stock return and uncertainty when in regime $s_t = 1$, this is seen from the wider upper and lower 5th percentiles. We notice that Excess Bond Returns $corp_t$ load very strongly for regime $s_t = 1$ with a coefficient of 0.388. Such a positive loading suggest that if Excess Bond Returns increase, Excess Stock Returns must also increase in the next period. However, the loading is reduced to 0.139 when conditioned on regime $s_t = 2$. We also see that the Dividend-to-Price Ratio dp_t and log Stock-Variance sv_t load positively on Excess Stock Returns for both regimes. In general we can say that all coefficients seem to be roughly 2-3 times smaller in regime $s_t = 2$ than the coefficients in regime $s_t = 1$. We also see that the coefficients also have a 2-3 larger dispersion in regime $s_t = 1$ than in regime $s_t = 2$.

A strong shift in the posterior distribution of Excess Bond Return predictors is also present. The intercept of Excess Bond Returns shows a tenfold shift when switching from regime $s_t = 1$ to regime $s_t = 2$. The posterior std. dev of the intercept is about 3 times larger for regime $s_t = 1$ than in regime $s_t = 2$ but the lower 5th percentile is roughly the same as opposed to the higher 5th percentile. This suggests more skewness in Excess Bond Returns during regime $s_t = 1$ than for regime $s_t = 2$. Excess Bond Returns are not highly auto-regressive as opposed to bond-yields, but there is still some predictability in

Table 2: This table shows the posterior distribution of parameters for the MS-VAR model. Each response variable is show separately per panel. From this posterior distribution we present the mean, median, standard deviation and the HPD estimates of the 5th and 95th percentiles. Each panel represents a VAR-equation with the response variable in the header and the predictor variables per single row.

MS-VAR Coefficient Posterior Distribution

Coeff.	Post. mean	Post. median	Post. std	5th perc.	95th perc.
		r_{t+1}			
$const_t\|s_t = 1$	8.897	8.916	5.419	-1.522	19.639
$const_t\|s_t = 2$	4.005	3.996	2.403	-0.727	8.622
$corp_t\|s_t = 1$	0.388	0.387	0.141	0.117	0.667
$corp_t\|s_t = 2$	0.139	0.139	0.079	-0.012	0.300
$dp_t\|s_t = 1$	1.738	1.737	1.127	-0.460	3.977
$dp_t\|s_t = 2$	0.412	0.412	0.416	-0.397	1.226
$sv_t\|s_t = 1$	0.613	0.618	0.512	-0.397	1.609
$sv_t\|s_t = 2$	0.252	0.251	0.242	-0.213	0.733
		$corp_{t+1}$			
$const_t\|s_t = 1$	5.760	5.781	3.882	-1.644	13.600
$const_t\|s_t = 2$	0.506	0.514	1.321	-2.053	3.145
$corp_t\|s_t = 1$	0.068	0.068	0.093	-0.114	0.251
$corp_t\|s_t = 2$	0.155	0.155	0.044	0.069	0.242
$dp_t\|s_t = 1$	0.333	0.331	0.776	-1.156	1.877
$dp_t\|s_t = 2$	-0.015	-0.014	0.225	-0.456	0.423
$sv_t\|s_t = 1$	0.678	0.679	0.351	-0.027	1.353
$sv_t\|s_t = 2$	0.076	0.076	0.127	-0.176	0.322
		dp_{t+1}			
$const_t\|s_t = 1$	-0.155	-0.155	0.055	-0.263	-0.051
$const_t\|s_t = 2$	-0.048	-0.048	0.024	-0.096	-0.001
$corp_t\|s_t = 1$	-0.004	-0.004	0.001	-0.007	-0.001
$corp_t\|s_t = 2$	-0.001	-0.001	0.001	-0.003	0.000
$dp_t\|s_t = 1$	0.974	0.974	0.011	0.952	0.996
$dp_t\|s_t = 2$	0.996	0.996	0.004	0.988	1.004
$sv_t\|s_t = 1$	-0.012	-0.012	0.005	-0.022	-0.002
$sv_t\|s_t = 2$	-0.004	-0.004	0.002	-0.009	0.001
		sv_{t+1}			
$const_t\|s_t = 1$	-0.258	-0.262	0.270	-0.776	0.285
$const_t\|s_t = 2$	-0.009	-0.009	0.041	-0.091	0.072
$corp_t\|s_t = 1$	-0.005	-0.005	0.006	-0.017	0.006
$corp_t\|s_t = 2$	-0.001	-0.001	0.001	-0.003	0.002
$dp_t\|s_t = 1$	0.010	0.009	0.052	-0.090	0.115
$dp_t\|s_t = 2$	-0.002	-0.002	0.007	-0.015	0.012
$sv_t\|s_t = 1$	0.952	0.952	0.024	0.905	1.000
$sv_t\|s_t = 2$	1.000	1.000	0.004	0.992	1.007

lagged Excess Bond Returns $corp_t$. The loading is not very pronounced in regime $s_t = 1$ but it roughly doubles in regime $s_t = 2$. The posterior std. dev also halves when switching from regime $s_t = 1$ to regime $s_t = 2$ making a more robust predictor. The Dividend-to-Price ratio seems to be random due to the large variance around the coefficient estimate. The log Stock-Variance predictor loads highly in regime $s_t = 1$ but the effect seems to be spurious in regime $s_t = 2$.

The Dividend-to-Price ratio has a stronger decay rate in regime $s_t = 1$ as opposed to $s_t = 2$. The intercept of Dividend-to-Price is negative for both regimes, but more so in regime $s_t = 1$ than for regime $s_t = 2$. The posterior std. dev of the intercept is also 2 times larger for the regime $s_t = 1$ as opposed to regime $s_t = 2$. The loadings from Excess Bond Returns seem to be zero as the posterior distribution is strongly peaked around zero. The Dividend-to-Price ratio is highly auto-regressive as we can see that the loading is very close to 1 in regime $s_t = 2$, the 95th percentile even seems to be larger than 1 which may suggests an explosive process in some of the draws from the posterior distribution. However this is not the case for regime $s_t = 1$. The loadings from log Stock-Variance seem to be slightly negative for both regimes.

Stock-Variance has a stronger decay rate in regime $s_t = 1$ as opposed to regime $s_t = 2$. This indicates that stock variance is likely to change more rapidly in regime $s_t = 1$ as opposed to regime $s_t = 2$. The loadings from Dividend-to-Price ratio and Excess Bond Returns seem to be random for both regimes as they are strongly peaked around zero. The intercept for regime $s_t = 2$ is strongly peaked around zero, however for regime $s_t = 1$ the effect is much more random and slightly negative on average.

If we turn our attention to the posterior distribution of the covariance matrix in Table 3 we see that the variance of Excess Stock Returns $Var(r_t|s_t = 1) = 31.02$, which is 19.3% volatility on an annualized basis. The variance of stock returns is strongly reduced in regime 2 with an annualized volatility of 12.4%, this is 55% lower volatility than in regime 1. There is a strong regime-shift in the covariance between stock returns and the Dividend-to-Price ratio. The negative covariance becomes much less pronounced, which makes it less suitable in forecasting the long-run conditional variance. Excess Bond Returns have a variance of 12.6 during regime $s_t = 1$ and 3.8 during regime $s_t = 2$, which is an annualized volatility of 12.3% and 6.7% respectively. The covariance between excess stock and bond returns changes when switching regimes, the covariance doubles between stocks and bonds indicating a stronger joint process during regime $s_t = 1$. However we see a reduction in correlation terms from 0.27 in regime $s_t = 1$ to 0.19 in regime $s_t = 2$. We have estimated a model that jointly captures the volatility dynamics of the joint distribution of excess stock and bond returns. In light of these findings we are comfortable to label regime 1 as "Risk-On" state and regime 2 as "Risk-Off" state.

If we investigate the transition probabilities in Table 4 we see that the persistence probability of

Table 3: This table describes the posterior distribution of the covariance matrix for the MS-VAR model. From this posterior distribution we present the mean, median, standard deviation and the HPD estimates of the 5th and 95th percentiles. Parameters are sampled on a monthly basis with the sampling period starting from July-1952 until December 2013

MS-VAR Covariance Matrix Posterior Distribution

Coeff.	Post. mean	Post. median	Post. std	5th perc.	95th perc.
$Var(r_t\|s_t = 1)$	31.021	30.711	3.768	23.964	38.418
$Var(r_t\|s_t = 2)$	12.860	12.827	0.884	11.162	14.619
$Cov(corp_t, r_t\|s_t = 1)$	4.011	3.934	1.772	0.550	7.514
$Cov(corp_t, r_t\|s_t = 2)$	1.932	1.926	0.339	1.260	2.589
$Var(corp_t\|s_t = 1)$	13.750	13.600	1.740	10.526	17.273
$Var(corp_t\|s_t = 2)$	3.779	3.768	0.271	3.248	4.310
$Cov(dp_t, r_t\|s_t = 1)$	-0.306	-0.303	0.037	-0.381	-0.238
$Cov(dp_t, r_t\|s_t = 2)$	-0.128	-0.127	0.009	-0.145	-0.111
$Cov(dp_t, corp_t\|s_t = 1)$	-0.038	-0.037	0.018	-0.073	-0.004
$Cov(dp_t, corp_t\|s_t = 2)$	-0.020	-0.019	0.003	-0.026	-0.013
$Var(dp_t\|s_t = 1)$	0.003	0.003	0.000	0.002	0.004
$Var(dp_t\|s_t = 2)$	0.001	0.001	0.000	0.001	0.001
$Cov(sv_t, r_t\|s_t = 1)$	-0.547	-0.540	0.127	-0.806	-0.312
$Cov(sv_t, r_t\|s_t = 2)$	-0.022	-0.022	0.011	-0.043	-0.002
$Cov(sv_t, corp_t\|s_t = 1)$	0.074	0.072	0.075	-0.073	0.224
$Cov(sv_t, corp_t\|s_t = 2)$	-0.008	-0.008	0.005	-0.018	0.003
$Cov(sv_t, dp_t\|s_t = 1)$	0.005	0.005	0.001	0.003	0.008
$Cov(sv_t, dp_t\|s_t = 2)$	0.000	0.000	0.000	0.000	0.000
$Var(sv_t\|s_t = 1)$	0.063	0.062	0.008	0.048	0.079
$Var(sv_t\|s_t = 2)$	0.003	0.003	0.000	0.003	0.004

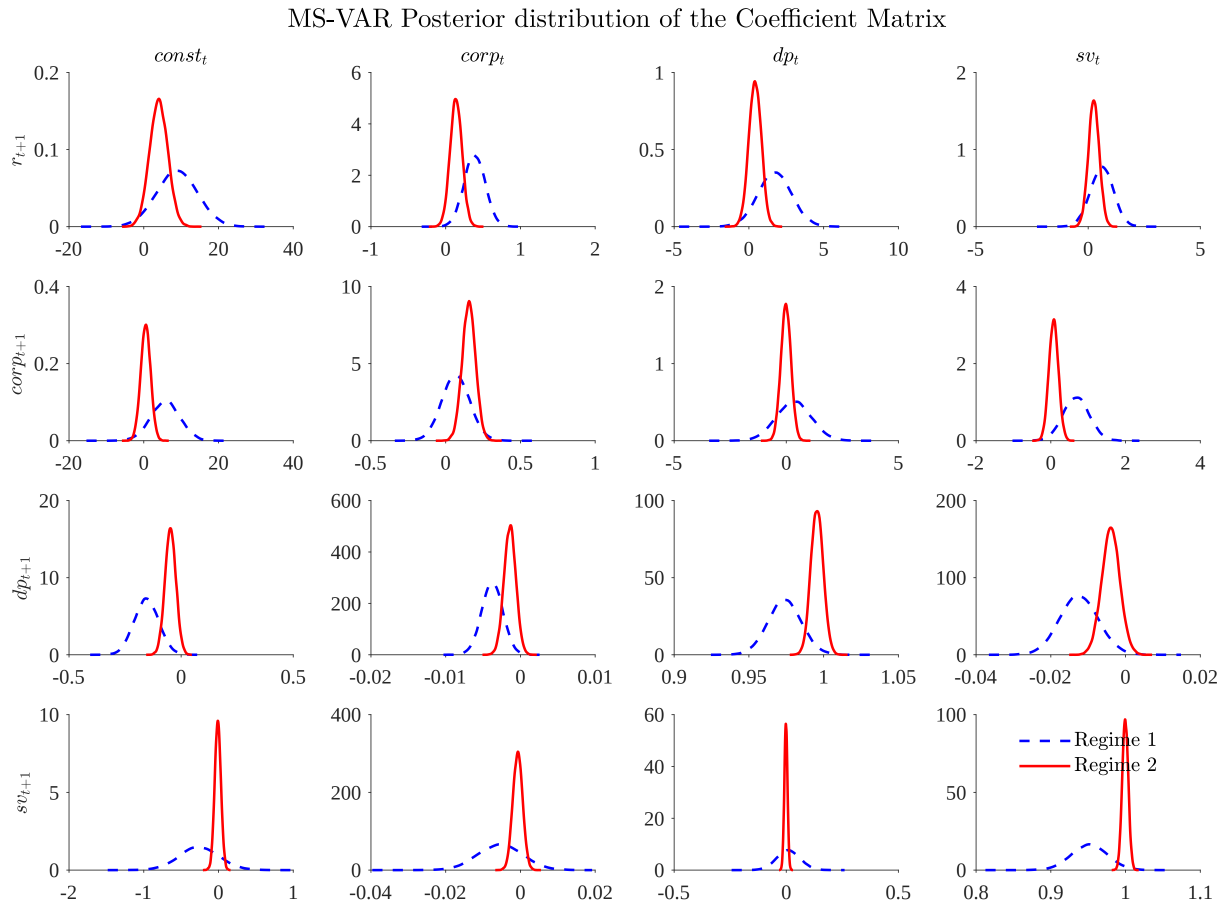

Figure 2: This figure shows the posterior distributions of parameters for the MS-VAR model. Each row represents a VAR-equation and each column represents a predictor variable. The labels of the each VAR-equation are denoted on the Y-axis and the predictor variables are labeled on the X-axis.

regimes are 69.4% and 90.9% respectively. The steady state probabilities are approximately 23.1% for regime $s_t = 1$ and 76.9% for the regime $s_t = 2$. The "Risk-On" regime lasts on average 3.3 months whereas the "Risk-Off" regime lasts 3 times longer for about 11.3 months on average. The average cumulative Excess Stock Return during the "Risk-On" regime is approximately -3.2% and 10% for the "Risk-Off" regime. The case for cumulative Excess Bond Returns during these regime is less clear as the average excess bond return during the "Risk-On" regime has tails of its posterior distribution in the negatives. If we ignore regimes we expect the excess stock return to be 0.46% per month. If we decompose the variance of excess stock returns, we can compute the variance explained by the model. The variance of the expectation of the variance is good, representing an explained variance approximation of 3.5% which is average for an Excess Return model on the general stock market.

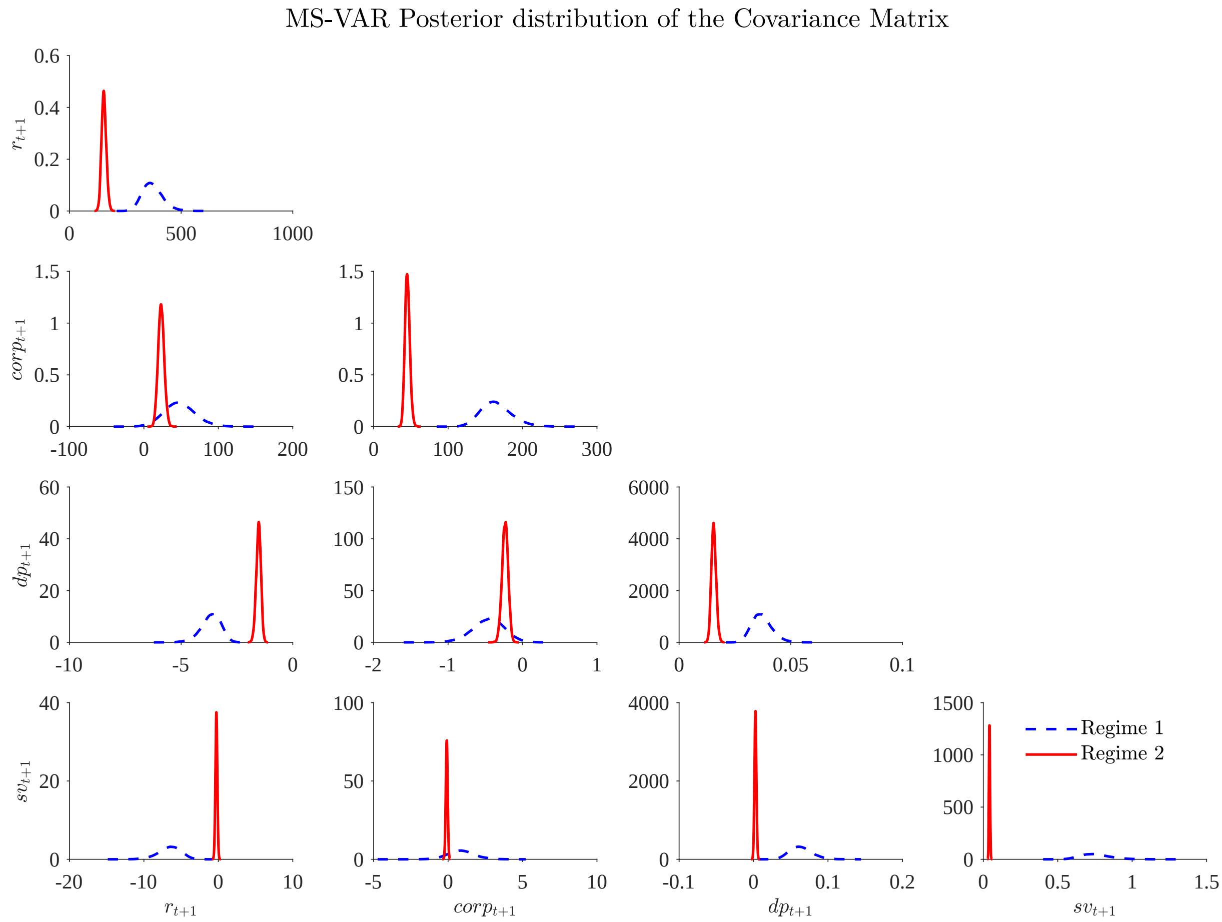

Figure 3: This figure shows the posterior distributions of covariance matrix for the MS-VAR model. The posterior density plots are shown for the lower triangular covariance matrix with labels on Y-Axis and X-Axis denoting the covariance between response variables, the diagonal elements are the variance density plots of each response variable.

8.2 Markov Switching VAR-TVTP Model

In this section we discuss the model results of the MS-VAR model extension with time-varying transition probabilities. In Table 7 the posterior distribution of the model parameters are presented, this table is accompanied by Figure 12 which includes the density plots of each model parameter. Table 8 shows the posterior distribution of the covariance matrix and is accompanied by Figure 13. The regime characteristics are shown in Table 6 and is accompanied by Figure 8. Additionally this model estimates the transition probability sensitivity to Stock Variance state variable, such that we consider predictability in the transition probability under uncertainty. Expanding the model to account for predictability of the transition probabilities does not alter the posterior distributions of the model parameters, therefore we

Table 4: This table shows the posterior distribution of transition probabilities, unconditional state probabilities and regime characteristics for the MS-VAR model. From this posterior distribution we present the mean, median, standard deviation and the HPD estimates of the 5th and 95th percentiles. $E(D|s_t)$ represents the average regime duration, $E(\sum r_t|s_t)$ is the average cumulative return during regime s_t.

Coeff.	Post. mean	Post. median	Post. std	5th perc.	95th perc.
Transition probabilities					
$p_{1,1}$	0.694	0.695	0.047	0.599	0.781
$p_{1,2}$	0.306	0.305	0.047	0.219	0.401
$p_{2,1}$	0.091	0.090	0.015	0.062	0.121
$p_{2,2}$	0.909	0.910	0.015	0.879	0.938
Unconditional state probabilities					
π_1	0.231	0.230	0.036	0.163	0.301
π_2	0.769	0.770	0.036	0.699	0.837
Regime characteristics					
$E(D\|s_t = 1)$	3.343	3.281	0.531	2.372	4.394
$E(D\|s_t = 2)$	11.260	11.052	1.917	7.736	15.014
$E(\sum r_t\|s_t = 1)$	-3.210	-3.160	1.637	-6.616	-0.168
$E(\sum r_t\|s_t = 2)$	10.052	9.825	2.436	5.728	15.049
$E(\sum corp_t\|s_t = 1)$	1.317	1.293	1.048	-0.729	3.395
$E(\sum corp_t\|s_t = 2)$	0.837	0.810	1.008	-1.161	2.816
$E(r_t\|s_t)$	0.465	0.465	0.165	0.142	0.787
$E(Var(r_t\|s_t)\|s_t)$	17.880	17.812	1.253	15.558	20.402
$Var(E(r_t\|s_t)\|s_t)$	0.661	0.608	0.360	0.059	1.373
$Var(r_t\|s_t)$	18.542	18.460	1.312	16.120	21.186
$\frac{Var(E(r_t\|s_t)\|s_t)}{Var(r_t\|s_t)}$	0.035	0.033	0.019	0.004	0.072

will not discuss these nor the covariance matrix. The regime dynamics are slightly altered which we will discuss more in depth. Tables 7, 8 and Figures 12 and 13 are provided in the appendix for reference, any differences with the MS-VAR model discussed in the previous section are simply due to sampling error.

Table 5: This table shows the Log Marginal Likelihoods for the MS-VAR and MS-TVTP models. We consider a 4th predictor in the VAR-equation next to Excess Stock Returns, Excess Bond Returns and Dividend-to-Price. This 4th predictor was then also used as a predictor for the transition probabilities in the MS-TVTP model. Bayes Factors are computed only for the MS-VAR vs MS-VAR-TVTP model comparison as they are non-nested.

Log Marginal Likelihoods

State variable	MS-VAR	MS-VAR-TVTP	Bayes Factor
Stock Variance	-687.85	-1089.89	402.04
Default Yield Spread	922.16	1065.75	-143.59
Default Return Spread	-1233.44	-1439.45	206.00
Term Spread	556.39	402.26	154.13
Inflation	-1039.18	-1339.14	299.96

MS-VAR Posterior distribution of the Regime Characteristics

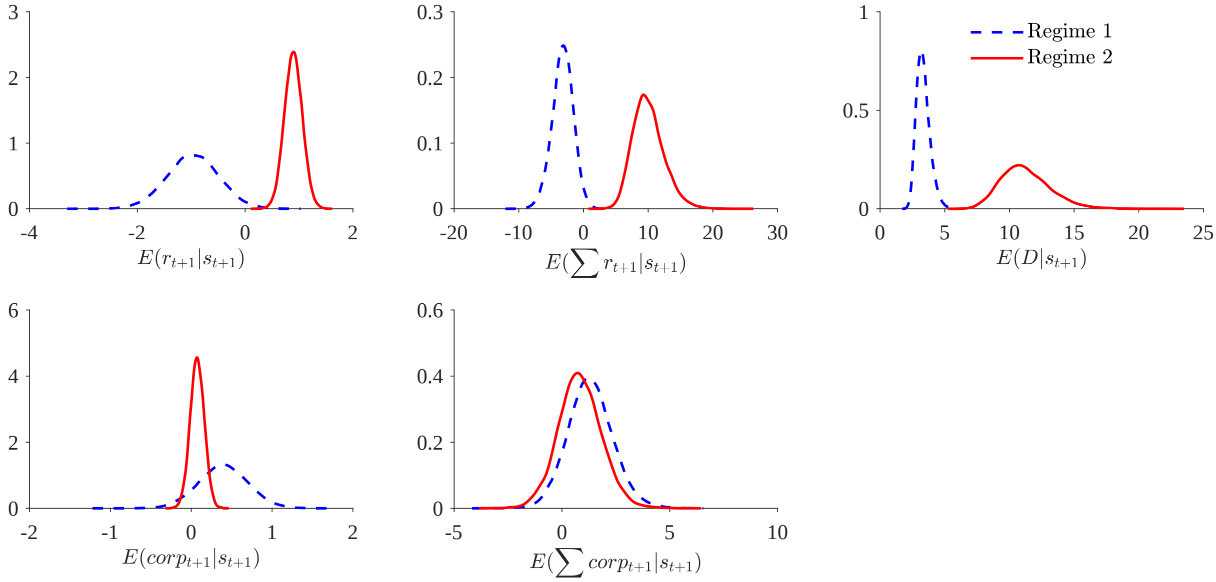

Figure 4: This figure shows the posterior distributions of stock and bond returns for both regime from the MS-VAR model. $E(D|s_t)$ represents the average regime duration, $E(r_t|s_t)$ is the average return during regime s_t. $E(\sum r_t|s_t)$ is the average cumulative return during regime s_t and is computed as the product between the average return and regime duration.

Table 6 shows the posterior distribution of the transition probabilities coefficients. The coefficients $\gamma^z_{1,2}$ and $\gamma^z_{2,2}$ indicate the sensitivity of the state variable z_t to the transition probability of staying in the same regime, i.e the regime persistence probability. We estimated both the time-varying component and the time-invariant average state persistence from equation (3.9). Since we have a reference state/regime we only need to estimate the transition probability from the first state to the second state, secondly we need to estimate the persistence probability for the second state. The model specification is dependent on the current state and thus it would grow very large if you add more states as the transition matrix grows quadratically with each state. Looking at the time-varying components $\gamma^Z_{1,2}$ and $\gamma^Z_{2,2}$ we see negative loadings meaning a increasing effect for increasing stock variance, however if we look to the time-invariant coefficients $\gamma_{1,2}$ and $\gamma_{2,2}$ we see that the coefficient that governs the transition probability is negative, but the the coefficient that governs the persistence probability is positive. This means that starting from state 1, the time-invariant coefficient tends to increase the probability of moving out of the state, but starting from state 2, the time-invariant coefficient tends to increase the persistence of the state. Looking at Figure 5 we see the Stock Variance regime effect. On the Y-axis we see the probability of remaining in the current state for the next period, on the X-axis we see the distribution of log Stock Variance. As log Stock-Variance increases, we see the persistence of regime 1 increases likewise, suggesting that large

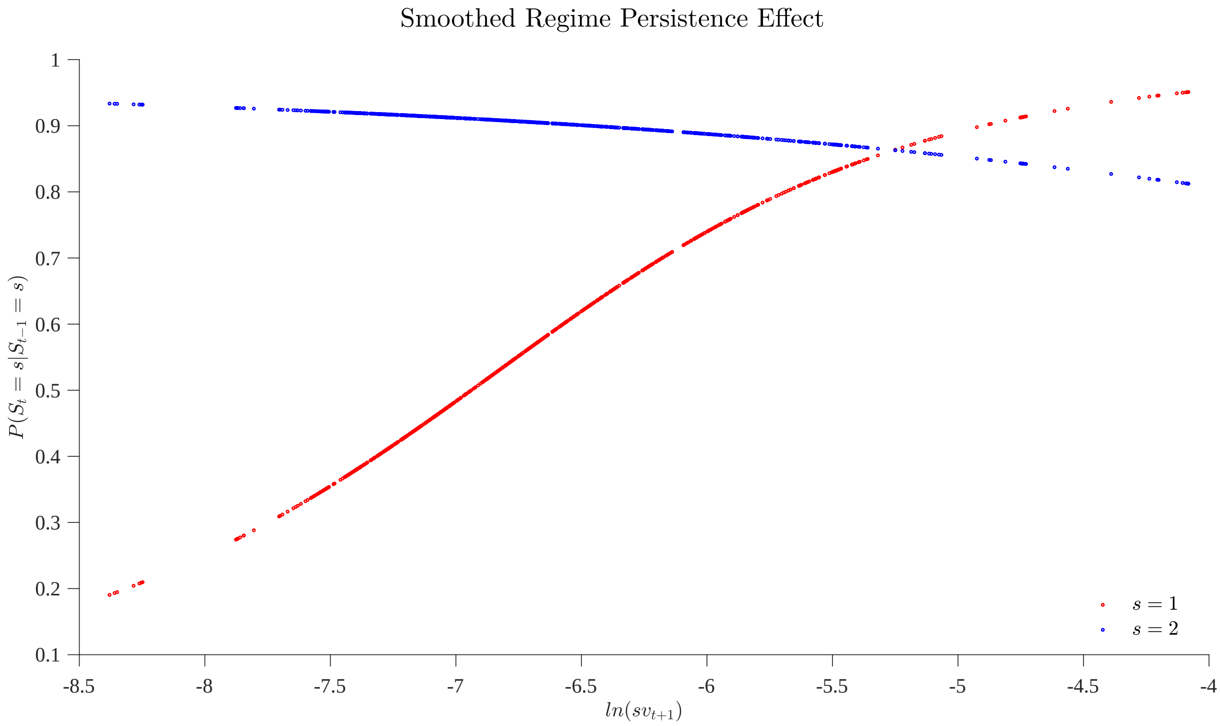

Figure 5: This figure shows the smoothed transition probability conditioned on log Stock-Variance for the MS-VAR-TVTP Model. On the Y-axis we see the probability of remaining in the current state for the next period and the X-axis denotes the log Stock-Variance.

Stock Variances is characteristic of the ”Risk-On” regime and thus increasing this regime persistence. Consequently we see the regime 2 ”Risk-Off” persistence to decrease on increasing Stock Variance. These result suggest that Stock Variance is a reasonable measure of regime persistence. Other measures that can be seen as a proxies for risk such as the term-spread, default-yield-spread, default-return-spread or inflation did not exhibit regime persistence predictability (results shown in Table 5), their smoothed regime persistence effect was much more flat. A flat regime persistence effect indicates that the model cannot separate the time varying from the invariant component, indicating no predictability.

Figure 7 shows the smoothed probability of the ”Risk-On” Regime for the time-varying model on right Y-axis and the total Excess Stock Return on the left Y-axis over the entire sample period. The probability fluctuates between zero and hundred percent and is consistent with an average short durations of 3 months for the ”Risk-On” regime. We can clearly see that the model captures the crash of 1987. Prior to this crash we saw positive increasing equity prices at relatively low volatility. For the ’dot-com’ or ’TMT’ bubble of early 2000s’ we saw an increase in equity prices at very low levels of volatility again and the same holds for the global financial crisis of 2008. For earlier periods we see similar patterns, but these are much less pronounced.

Table 6: This table shows the posterior distribution of transition probabilities, unconditional state probabilities and regime characteristics for the MS-VAR-TVTP model. From this posterior distribution we present the mean, median, standard deviation and the posterior density estimates of the 5th and 95th percentiles. $E(D|s_t)$ represents the average regime duration, $E(\sum r_t|s_t)$ is the average cumulative return during regime s_t and is computed as the product between the average return and regime duration. $\gamma^z_{k,2}$ denotes the transition probability coefficient and $\gamma_{k,2}$ is the time-invariant component.

Coeff.	Post. mean	Post. median	Post. std	5th perc.	95th perc.
Transition probabilities					
$\gamma^z_{1,2}$	-1.129	-1.109	0.351	-1.824	-0.459
$\gamma_{1,2}$	-0.425	-0.432	0.276	-0.970	0.123
$\gamma^z_{2,2}$	-0.262	-0.259	0.232	-0.738	0.169
$\gamma_{2,2}$	2.240	2.235	0.182	1.896	2.606
Regime characteristics					
$E(D|s_t = 1)$	3.067	3.054	0.326	2.429	3.696
$E(D|s_t = 2)$	10.515	10.491	1.091	8.446	12.705
$E(\sum r_t|s_t = 1)$	-2.984	-2.978	1.486	-5.919	-0.140
$E(\sum r_t|s_t = 2)$	9.289	9.222	1.817	5.732	12.828
$E(\sum corp_t|s_t = 1)$	1.278	1.273	0.964	-0.558	3.289
$E(\sum corp_t|s_t = 2)$	0.740	0.728	0.921	-1.044	2.576
$E(r_t|s_t)$	0.467	0.468	0.152	0.168	0.762
$E(Var(r_t|s_t)|s_t)$	17.871	17.824	1.085	15.809	20.040
$Var(E(r_t|s_t)|s_t)$	0.655	0.605	0.356	0.063	1.356
$Var(r_t|s_t)$	18.526	18.474	1.116	16.310	20.660
$\frac{Var(E(r_t|s_t)|s_t)}{Var(r_t|s_t)}$	0.035	0.033	0.019	0.003	0.071

Figure 6 shows the smoothed transition probability over time obtained from the MS-VAR-TVTP model. We can see substantial fluctuation over the sample period with lows of 30% during early '60s for the persistence of the "Risk-On" regime. We also see several large swings in transition and persistence probabilities during global equity market downturns, which is in line with the models' intended behavior.

Taking both these models in consideration we conclude that there is sufficient evidence to say that temporal parameter shifts are present in the economic processes analyzed and that stock variance may be a good predictor for the transition probability between regimes.

8.3 Conditional Variance

The key difference between the MS-VAR and MS-VAR-TVTP model is the predictability of the transition state probability versus that of a constant assumed transition probability. Using the forecasted returns from each Gibbs sample we are able to compute the conditional variance over each sample draw for each investment horizon. Naturally we would expect the conditional variance to increase as we increase the forecasting horizon for a model with no predictability with *i.i.d.* stock returns and incorporating

Smoothed Transition Probabilities over Time

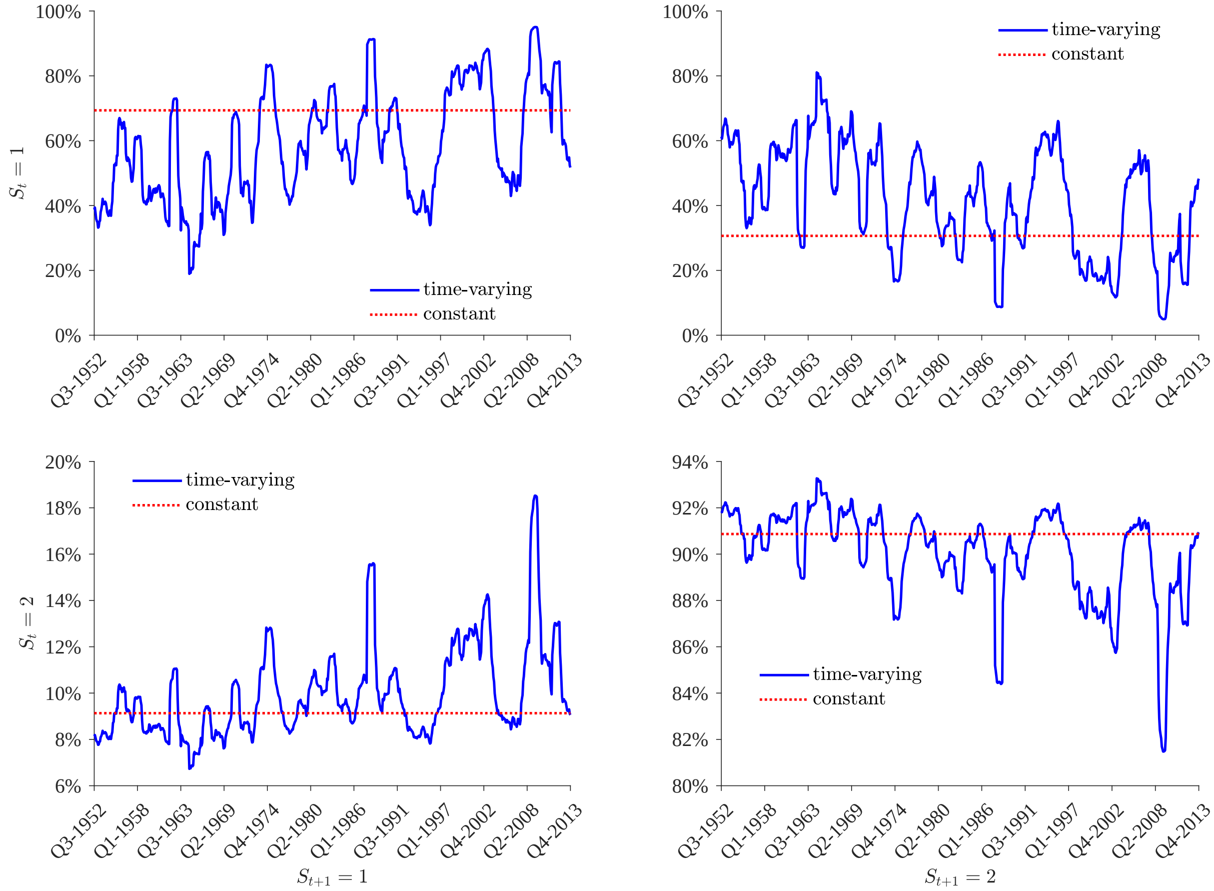

Figure 6: This figure shows the smoothed time-varying transition probabilities for the MS-VAR-TVTP modoel. The constant transition probabilities from MS-VAR model have been added for comparison. The models are denoted as 'time-varying' and 'constant' respectively. Each quadrant represents an element from the transition probability matrix from equation (3.1).

parameter uncertainty. In models that do not account for regime uncertainty we see that long horizon conditional return variances tend to decrease Brandt [2010] and Barberis [2000]. Figure 9 depicts the conditional variance of forecasted excess returns. We forecast 120 Months in to the future. The key goal of this exercise is to investigate whether the conditional variance decreases over longer horizons. Decreasing conditional variances over the longer horizon indicates a reduction in long term risk of the asset class, therefore making it more attractive for long-term investors. This is the so-called "horizon effect", see Barberis [2000]. The top panel depicts the conditional variance of stock returns and the bottom panel shows the conditional variance of corporate bond returns. Both panels make a comparison between the MS-VAR and MS-VAR-TVTP model forecasted returns up to an investment horizon of 120 months.

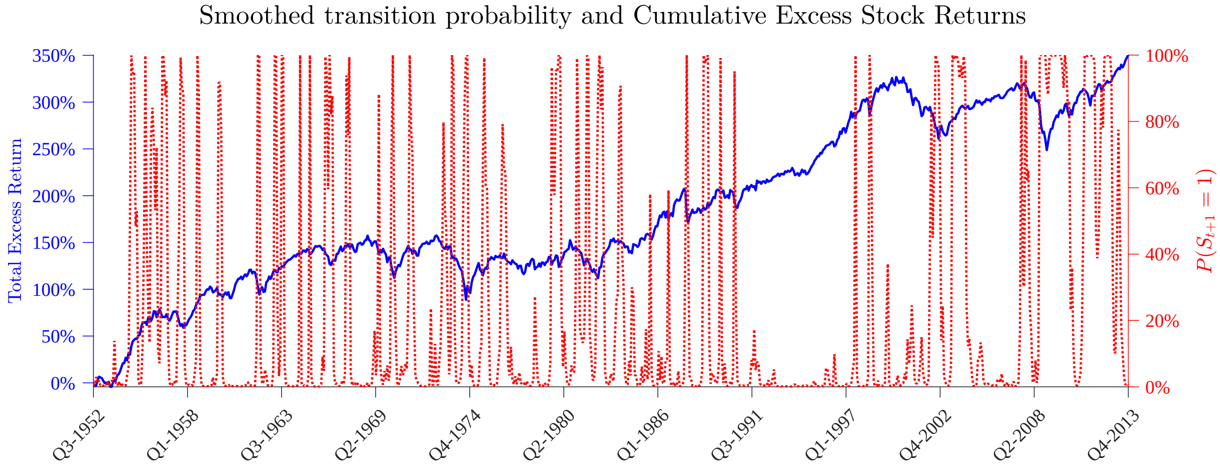

Figure 7: This figure shows the smoothed state probability for $S_t = 1$ for both models overlayed with the total excess return of the stock market. The Y-axis denotes the smoothed probability and the X-axis runs from July-1952 until December 2013.

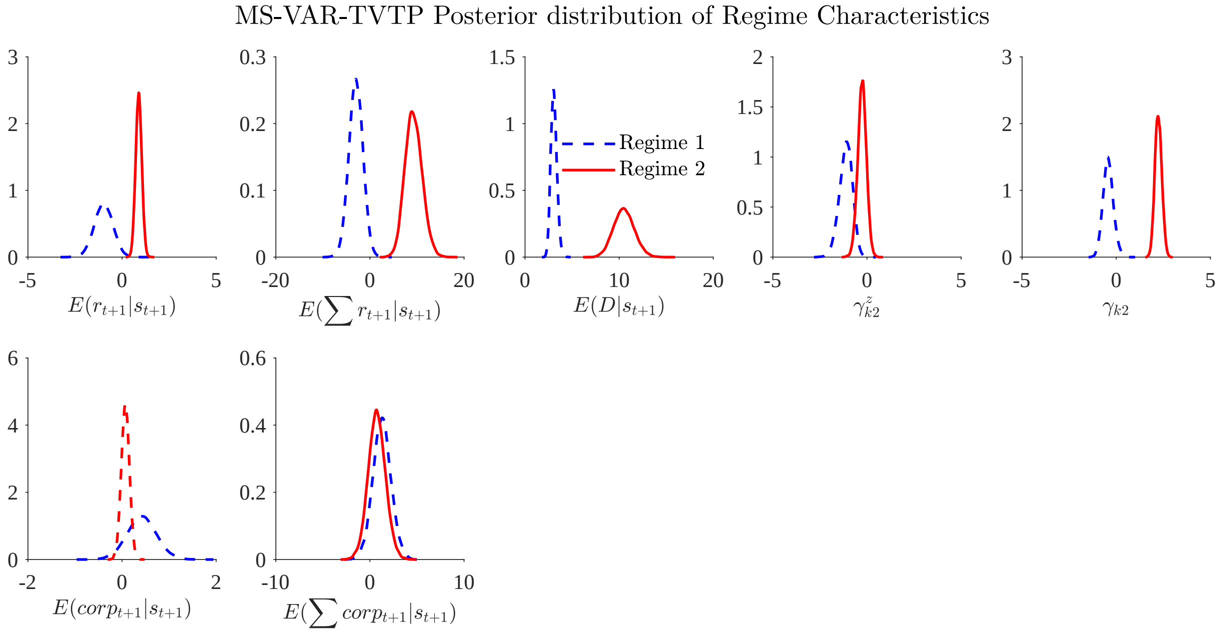

Figure 8: This figure shows the posterior distribution of regime characteristics for the MS-VAR-TVTP model conditioned on both regimes. $E(D|s_t)$ represents the average regime duration, $E(r_t|s_t)$ is the average return during regime s_t. $E(\sum r_t|s_t)$ is the average cumulative return during regime s_t and is computed as the product between the average return and regime duration. $\gamma^z_{k,2}$ denotes the transition probability coefficient and $\gamma_{k,2}$ is the time-invariant component.

The conditional variance of the cumulative Excess Stock is reduced by about 10% when accounting for transition probability predictability. For both models, we see increasing conditional variances up to

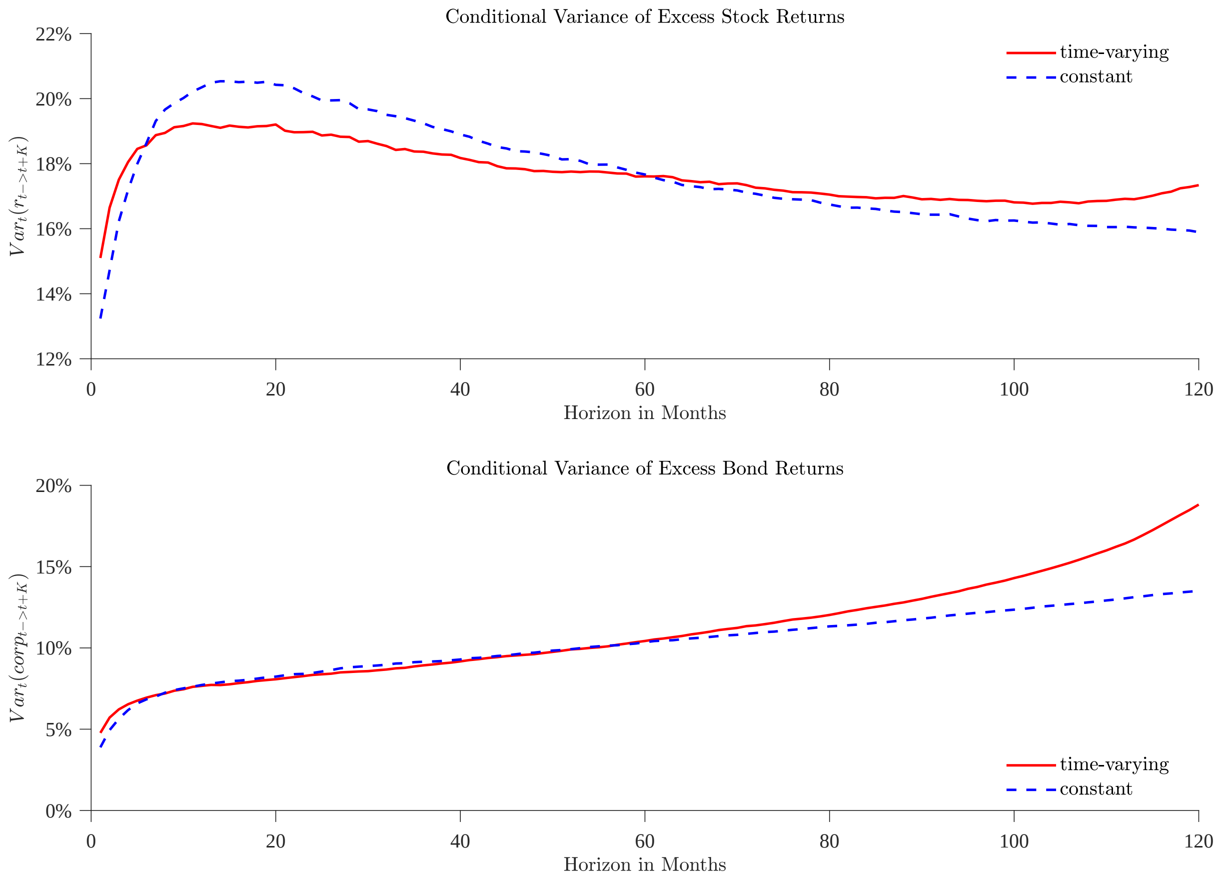

Figure 9: This figure shows the smoothed conditional variance for simulated returns for the both models. The MS-VAR model result is denoted as 'constant' and the TVTP model extension is denoted as 'time-varying'. Simulated returns are sampled from each Gibbs run.

horizon of 20 months, where the time-varying model conditional variance is below that of the constant model. These variances start to decrease from month 20 and decrease until month 60, where we see a divergence between time-varying and constant transition probabilities. After the 60th month the effect of transition probability predictability is reduced as it starts to generate slightly higher conditional variances and even tilts upwards after month 100. The constant transition probability model has sees its conditional variance reducing. There is no predictability effect for Corporate Bonds for longer investment horizons. The conditional variance follows roughly the same pattern until month 60 for both models. After the 60th month the conditional variances for bond returns are higher than those obtained from the constant transition probability model. These result imply that the model with transition probability predictability starts to deteriorate after month 60 for stocks and bonds.

8.4 Long Term Buy & Hold Portfolios

In this section we discuss the optimal Long Term Buy & Hold Portfolios that are derived using the MS-VAR and MS-VAR-TVTP Models. During each Gibbs sample-draw we simulate 120 Months of Excess Stock and Bond Returns, Stock Variance and Dividend-to-Price ratio. Using these simulations and predictability we are able to derive the optimal weights when investing with a so called Buy & Hold investment strategy. The optimal weights are obtained by maximizing the power utility function (7.3). In this strategy, we cannot alter our weights during investment period such that the asset weights may fluctuate during the investment period. Therefore the problem reduces to that of a myopic investor with single period if we consider the cumulative return for the entire holding period as a single period return. We investigate four different levels of risk aversion commonly used in academic research. The levels considered are $\gamma = \{2, 5, 10, 20\}$, the higher the level, the higher the risk aversion and thus a lower risk appetite. An investor with a high level of risk aversion will be more averse to large asset return volatility and will prefer lower returns at lower risk. Figure 10 shows the optimal capital allocation to stocks, bonds and the risk-free rate considered as a function of investment horizon in months. The risk-free rate is assumed to have a zero return, thus equal to a cash investment. Additionally the capital must be fully invested, such that a budget constraint is introduced for the allocation weights to sum to one. We used the MS-VAR and MS-VAR-TVTP models as described in the previous sections, denoted as constant and time-varying respectively. We see that the allocation to stocks rapidly decreases until the 20th month horizon for both models at all levels of risk aversion. The decline for the time-varying model is much less than for the constant model due to the lower conditional variances mentioned in the previous section. After the 20th month the allocation to stocks increases upwards, however at risk aversion level of $\gamma = 2$ the allocation is higher for the constant model, indicating the preference for a lower conditional variance at investment horizon levels larger than roughly 80 months. For risk aversion levels $\gamma \geq 5$ the optimal allocation for stocks converges to same allocation at an investment horizon of 10 years. The value added is clear from the time-varying model, investors with horizons of 6-7 years with moderate and higher risk aversion levels should account for time-varying transition probabilities. Buy-and-hold investors with high risk tolerance and investment horizons longer than 7 years can safely ignore time-varying transition probabilities. The allocation to bonds is zero until the 20th month for both models. After month 20, the allocation for both models with risk aversion levels of $\gamma \geq 5$ grow logarithmically, but the time-varying model show much higher levels of allocation to bonds. For $\gamma = 2$ we see roughly the same picture but the allocation actually reduces after month 50 and month 80 for time-varying and constant respectively.

The cash-investment rapidly increases only then to decrease after month 20 for all risk aversion levels. One note here is that time-varying model allocates less to cash as opposed to the constant model. The

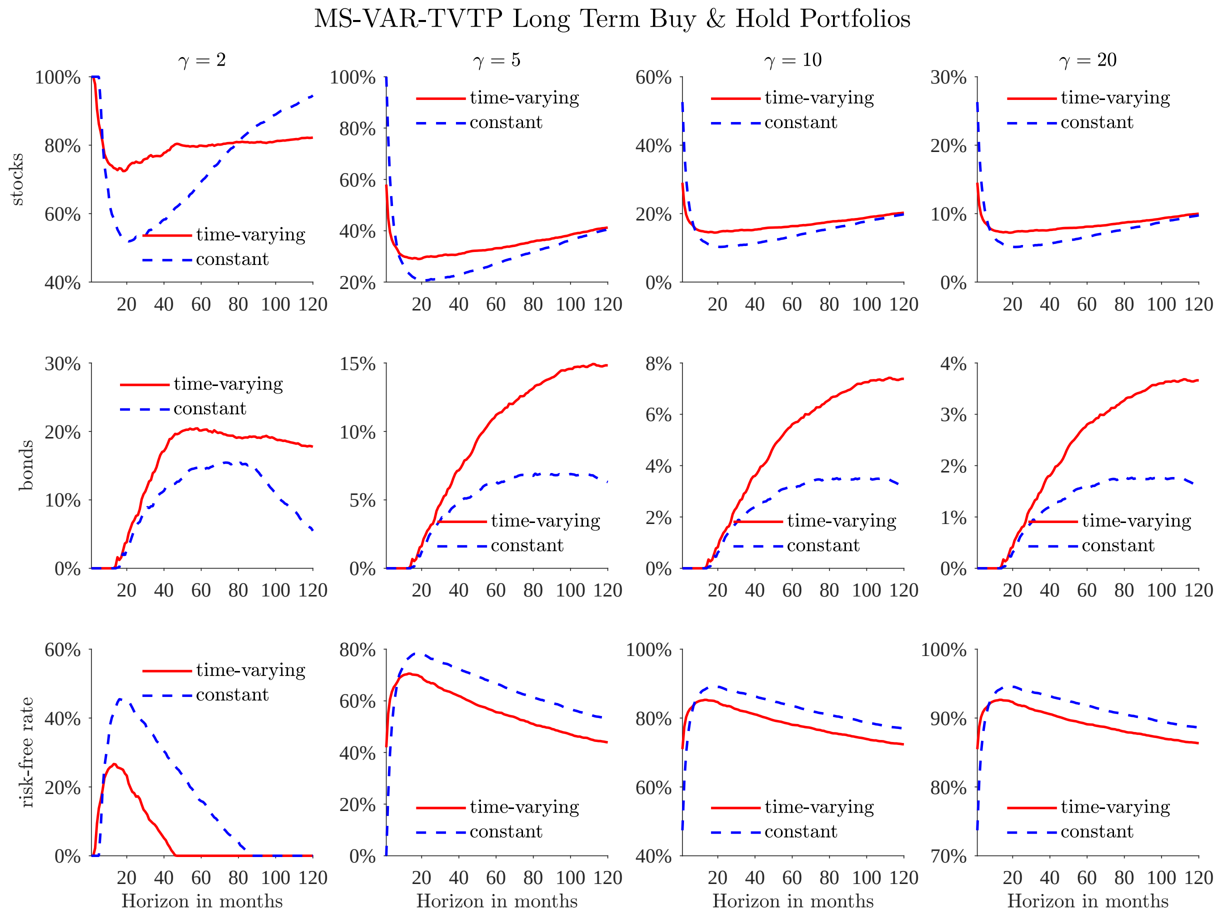

Figure 10: This figure shows MS-VAR-TVTP Buy & Hold Long Term Portfolio Weights. The Y-axis represents the optimal proportion invested in three asset classes, Stocks, Bonds and Risk-free rate. The X-axis shows the investment horizon in months, where 10 years at most is considered.

optimal allocation for a buy-and-hold investor with a horizon of 5 years and very low risk aversion level is approximately 80% invested in the stock market and 20% in corporate bonds. The allocation of an asset class is inversely controlled by the scaling factor of the step size in γ i.e, let K be a real number, then K_n/K_{n-1}, where $K_n > K_{n-1}$ save for measurement errors. This means we can deduce approximately the optimal allocation for any risk aversion level from another risk aversion level. Plugging in the $\gamma = 10$ and assuming the optimal stock allocation at $\gamma = 2$ equals 80% at a horizon of 80 months, means that the optimal investment approximately is roughly $2/10 \times 80\% = 16\%$ at a horizon of 80 months. This is consistent with the derived optimal weights from Figure 10.

Smoothed Certainty Equivalent Returns

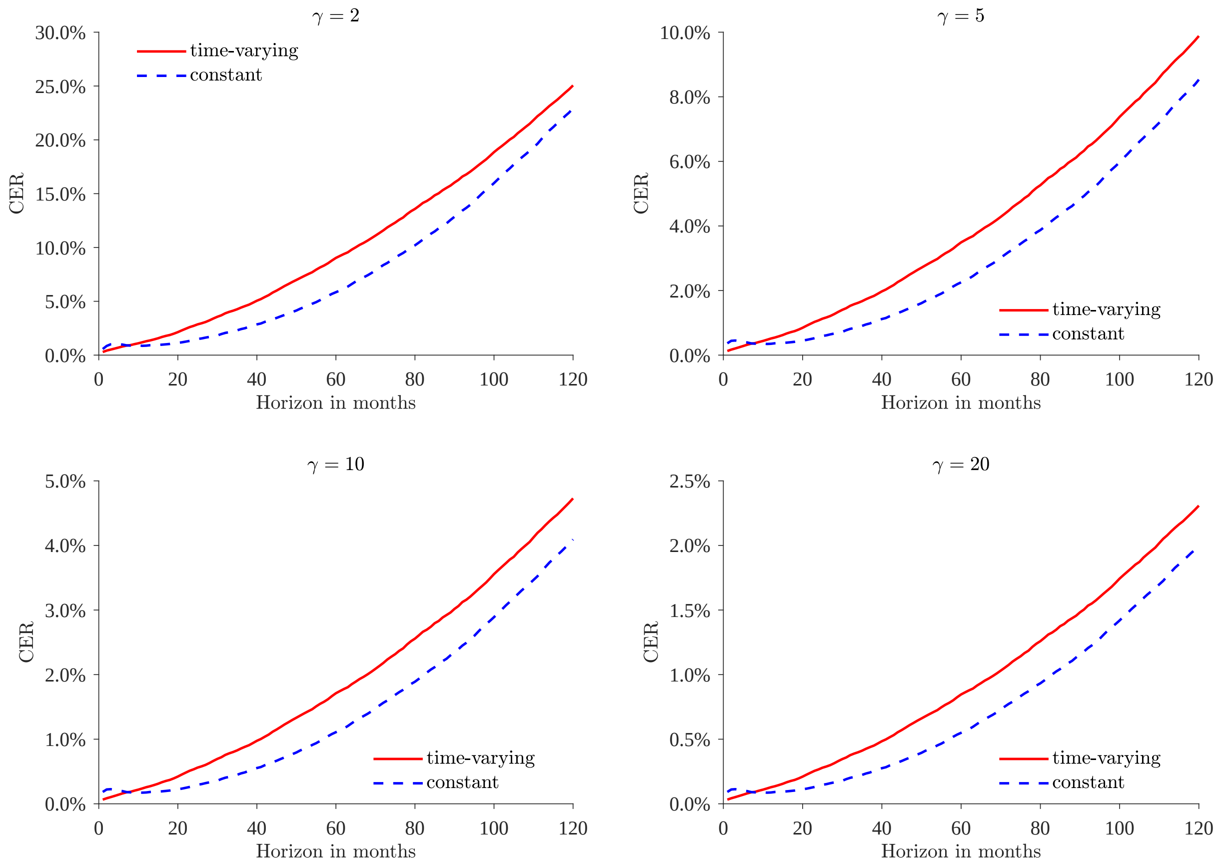

Figure 11: This figure shows the certainty equivalent returns for four levels of risk aversion. Returns from each Gibbs sample have used to estimate the utility function.

8.5 Certainty Equivalent Return

The certainty equivalent return (CER) is the level of return where the investor is indifferent between any more risk or return, thus being the optimal level of portfolio return. Figure 11 shows the CER as a function of the investment horizon. In each quadrant the CER of MS-VAR and the MS-VAR-TVTP models are considered for comparison. We consider four levels of risk aversion, $\gamma = 2$ which represents Low Risk, $\gamma = 5$ represents Medium Risk, $\gamma = 10$ represents High Risk and $\gamma = 20$ for Very High Risk.

Consistent with the an increasing level of risk aversion we see the scale of CER decreasing. We notice that the CER is higher for the MS-VAR-TVTP model for each level of risk aversion starting from an investment horizon of 10 months. This indicates that the MS-VAR-TVTP model is able to produce a higher level of economic utility than the MS-VAR model up at all investment horizons. This suggests that the predictability of regime transition probability increases utility for the buy-and-hold investor at all levels of risk aversions and investment horizons longer than 1 year.

9 Conclusion

This Master thesis is on the application of Markov Switching Vector Autoregressive Models (MS-VAR) and the predictability of long term strategic asset allocation under uncertainty. The main contribution of this thesis is the extension of such models with Time Varying Transition Probabilities (TVTP). The literature in this field currently does not provide a clear and intuitive way to incorporate predictability for future economic or market states for these types of models that incorporate condition on previous regimes. There is a strong regime-shift effect in the predictor set of Excess Stock Returns. We separate two economic regimes by permuting on the estimated excess stock return variance, where the first regime has a 55% higher level of stock market volatility relative to the second regime. We can therefore label these market regimes as ”Risk-On” and ”Risk-Off” market states. Incorporating time-varying transition probabilities reduces the conditional variance of simulated returns by a meaningful amount allowing long-term buy-and-hold investors benefit from regime predictability. By accounting for predictability for regimes we find that these models increase certainty equivalent returns (CER) for buy and hold investors with investment horizons of more than 1 year. This thesis contains a lot of decisions that still need thorough investigations, for example different regimes could possibly be identified when permuting on other parameters, such as covariances, correlations or coefficients. Changes in the model specification can impact the ability of the model to properly identify regimes, more so than parameter permutation. Bayesian averaging can be used to alleviate these research questions, but out of the scope for this master thesis. More lags or different lags could also be considered but severely increases model complexity and the number of parameters increases quadratically. The question of how much actual regimes we can identify is still unanswered, Pettenuzzo and Timmermann [2011] propose a more general model that incorporates structural breaks and use it to investigate the implications for portfolio choice. Furthermore other risk factors can be used to investigate predictability for regimes, such as liquidity or return factors such as value and momentum. These return factor are becoming more popular in the context of portfolio choice and asset allocators are wondering whether to tilt their portfolios to such return factors.

References

Andrew Ang and Geert Bekaert. International asset allocation with regime shifts. *The Review of Financial Studies*, 15(4):1137–1187, 2002.

Doron Avramov. Stock return predictability and model uncertainty. *Journal of Financial Economics*, 64 (3):423–458, 2002.

Nicholas Barberis. Investing for the long run when returns are predictable. *The Journal of Finance*, 55 (1):225–264, 2000.

Marco Bazzi, Francisco Blasques, Siem Jan Koopman, and Andre Lucas. Time-varying transition probabilities for markov regime switching models. *Journal of Time Series Analysis*, 38(3):458–478, 2017.

Michael W Brandt. Portfolio choice problems. In *Handbook of financial econometrics: Tools and techniques*, pages 269–336. Elsevier, 2010.

Gerhard Bry and Charlotte Boschan. Front matter to" cyclical analysis of time series: Selected procedures and computer programs". In *Cyclical Analysis of Time Series: Selected Procedures and Computer Programs*, pages 13–2. NBEr, 1971.

John Y Campbell and Robert J Shiller. Stock prices, earnings, and expected dividends. *The Journal of Finance*, 43(3):661–676, 1988.

John Y Campbell and Luis Viceira. *Strategic asset allocation: portfolio choice for long-term investors*. Clarendon Lectures in Economic, 2002.

Ming-Hui Chen and Qi-Man Shao. Monte carlo estimation of bayesian credible and hpd intervals. *Journal of Computational and Graphical Statistics*, 8(1):69–92, 1999.

Shiu-Sheng Chen. Predicting the bear stock market: Macroeconomic variables as leading indicators. *Journal of Banking & Finance*, 33(2):211–223, 2009.

Siddhartha Chib. Marginal likelihood from the gibbs output. *Journal of the American Statistical Association*, 90(432):1313–1321, 1995.

Siddhartha Chib. Calculating posterior distributions and modal estimates in markov mixture models. *Journal of Econometrics*, 75(1):79–97, 1996.

Drew Creal, Siem Jan Koopman, and André Lucas. Generalized autoregressive score models with applications. *Journal of Applied Econometrics*, 28(5):777–795, 2013.

Francis X Diebold, Joon-Haeng Lee, and Gretchen C Weinbach. Regime switching with time-varying transition probabilities. *Business Cycles: Durations, Dynamics, and Forecasting*, pages 144–165, 1994.

Eugene F Fama and Kenneth R French. Dividend yields and expected stock returns. *Journal of financial economics*, 22(1):3–25, 1988.

Andrew J Filardo. Business-cycle phases and their transitional dynamics. *Journal of Business & Economic Statistics*, 12(3):299–308, 1994.

Sylvia Frühwirth-Schnatter. *Finite mixture and Markov switching models*. Springer Science & Business Media, 2006.

Sylvia Frühwirth-Schnatter and Rudolf Frühwirth. Data augmentation and mcmc for binary and multinomial logit models. In *Statistical modelling and regression structures*, pages 111–132. Springer, 2010.

Edward Greenberg. *Introduction to Bayesian econometrics*. Cambridge University Press, 2012.

Massimo Guidolin and Allan Timmermann. Economic implications of bull and bear regimes in uk stock and bond returns*. *The Economic Journal*, 115(500):111–143, 2005.

Massimo Guidolin and Allan Timmermann. An econometric model of nonlinear dynamics in the joint distribution of stock and bond returns. *Journal of Applied Econometrics*, 21(1):1–22, 2006.

Massimo Guidolin and Allan Timmermann. Asset allocation under multivariate regime switching. *Journal of Economic Dynamics and Control*, 31(11):3503–3544, 2007.

James D Hamilton. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica: Journal of the Econometric Society*, pages 357–384, 1989.

James D Hamilton and Gang Lin. Stock market volatility and the business cycle. *Journal of Applied Econometrics*, 11(5):573–593, 1996.

James Douglas Hamilton. *Time series analysis*, volume 2. Princeton university press Princeton, 1994.

Chris C Holmes, Leonhard Held, et al. Bayesian auxiliary variable models for binary and multinomial regression. *Bayesian Analysis*, 1(1):145–168, 2006.

Shmuel Kandel and Robert F Stambaugh. On the predictability of stock returns: an asset-allocation perspective. *The Journal of Finance*, 51(2):385–424, 1996.

Sylvia Kaufmann. K-state switching models with time-varying transition distributions—does loan growth signal stronger effects of variables on inflation? *Journal of Econometrics*, 187(1):82–94, 2015.

Erik Kole and Dick van Dijk. How to identify and forecast bull and bear markets? *Journal of Applied Econometrics*, 32(1):120–139, 2017.

Hans-Martin Krolzig. *Markov-switching vector autoregressions: Modelling, statistical inference, and application to business cycle analysis*, volume 454. Springer Science & Business Media, 2013.

Martin Lettau and Stijn Van Nieuwerburgh. Reconciling the return predictability evidence. *Review of Financial Studies*, 21(4), 2008.

Asger Lunde and Allan Timmermann. Duration dependence in stock prices. *Journal of Business & Economic Statistics*, 22(3), 2004.

John M Maheu and Thomas H McCurdy. Identifying bull and bear markets in stock returns. *Journal of Business & Economic Statistics*, 18(1):100–112, 2000.

John M Maheu, Thomas H McCurdy, and Yong Song. Components of bull and bear markets: bull corrections and bear rallies. *Journal of Business & Economic Statistics*, 30(3):391–403, 2012.

Daniel McFadden et al. Conditional logit analysis of qualitative choice behavior. 1973.

Michael A Newton and Adrian E Raftery. Approximate bayesian inference with the weighted likelihood bootstrap. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 3–48, 1994.

Adrian R Pagan and Kirill A Sossounov. A simple framework for analysing bull and bear markets. *Journal of Applied Econometrics*, 18(1):23–46, 2003.

Davide Pettenuzzo and Allan Timmermann. Predictability of stock returns and asset allocation under structural breaks. *Journal of Econometrics*, 164(1):60–78, 2011.

Svetlozar T Rachev, John SJ Hsu, Biliana S Bagasheva, and Frank J Fabozzi. *Bayesian methods in finance*, volume 153. John Wiley & Sons, 2008.

Paul A Samuelson et al. Lifetime portfolio selection by dynamic stochastic programming. *The Review of Economics and Statistics*, 51(3):239–246, 1969.

Jules H van Binsbergen and Michael W Brandt. Solving dynamic portfolio choice problems by recursing on optimized portfolio weights or on the value function? *Computational Economics*, 29(3-4):355–367, 2007.

Dick van Dijk. Bayesian econometrics in finance: Week 5. Powerpoint Sheets, November 28 2014. Master Quantitative Finance.

Ivo Welch and Amit Goyal. A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, 21(4):1455–1508, 2008.

Tables & Figures

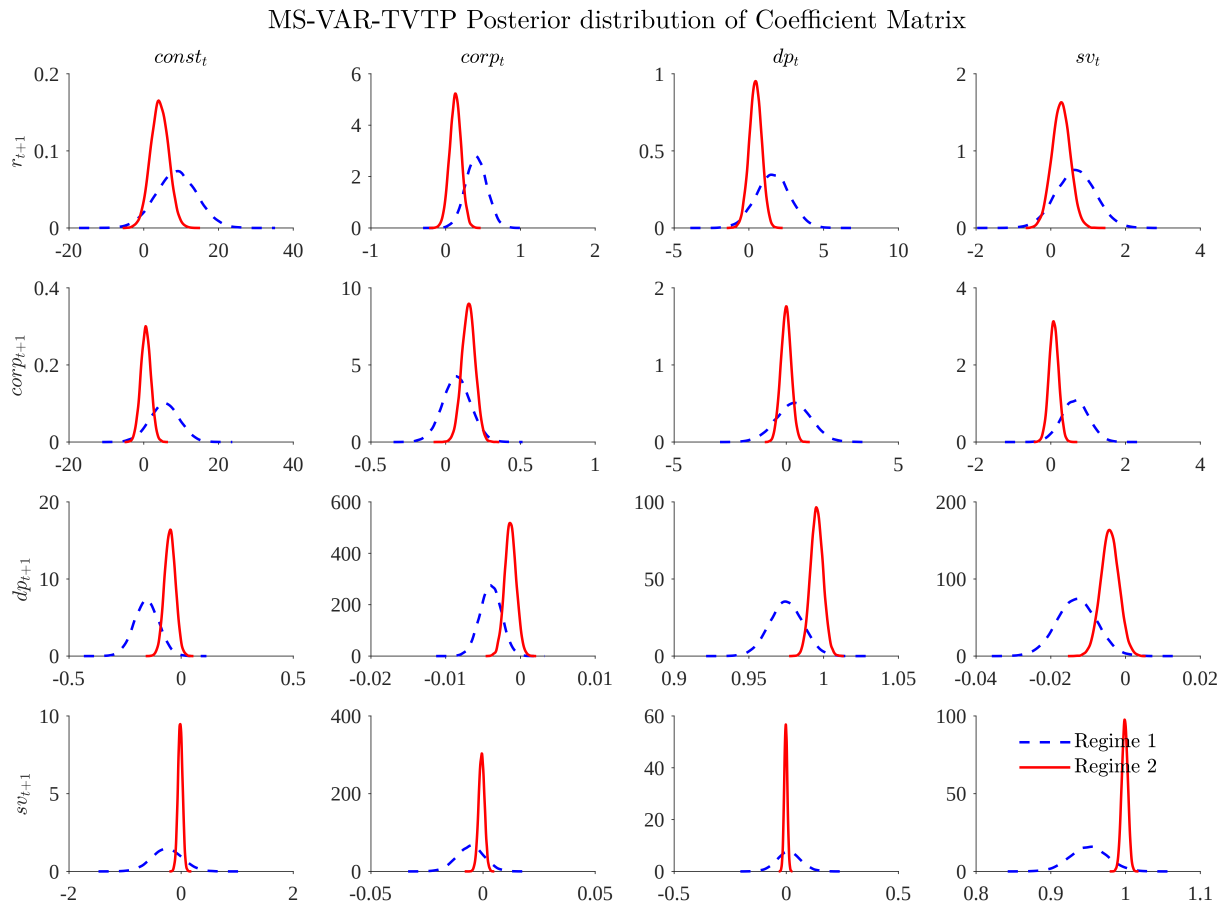

Figure 12: This figure shows the posterior distribution of parameters for the MS-TVTP model. Each row represents a VAR-equation and each column represents a predictor variable. The labels of the each VAR-equation are denoted on the Y-axis and the predictor variables are labeled on the X-axis.

Table 7: This table shows the posterior distribution of parameters for the MS-VAR-TVTP model. Each response variable is show separately per panel. From this posterior distribution we present the mean, median, standard deviation and the HPD estimates of the 5th and 95th percentiles. Each panel represents a VAR-equation with the response variable in the header and the predictor variables per single row.

MS-VAR-TVTP Coefficient Posterior Distribution

Coeff.	Post. mean	Post. median	Post. std	5th perc.	95th perc.
		r_{t+1}			
$const_t \vert s_t = 1$	8.736	8.750	5.430	-1.917	19.295
$const_t \vert s_t = 2$	4.218	4.198	2.392	-0.571	8.734
$corp_t \vert s_t = 1$	0.410	0.409	0.143	0.137	0.696
$corp_t \vert s_t = 2$	0.130	0.131	0.077	-0.020	0.284
$dp_t \vert s_t = 1$	1.612	1.614	1.127	-0.571	3.815
$dp_t \vert s_t = 2$	0.434	0.439	0.419	-0.389	1.248
$sv_t \vert s_t = 1$	0.667	0.668	0.526	-0.351	1.703
$sv_t \vert s_t = 2$	0.272	0.274	0.243	-0.195	0.758
		$corp_{t+1}$			
$const_t \vert s_t = 1$	5.591	5.589	3.944	-2.132	13.360
$const_t \vert s_t = 2$	0.538	0.537	1.345	-2.084	3.184
$corp_t \vert s_t = 1$	0.073	0.073	0.095	-0.113	0.261
$corp_t \vert s_t = 2$	0.153	0.153	0.044	0.066	0.238
$dp_t \vert s_t = 1$	0.328	0.336	0.790	-1.224	1.857
$dp_t \vert s_t = 2$	-0.012	-0.010	0.230	-0.479	0.423
$sv_t \vert s_t = 1$	0.654	0.656	0.361	-0.057	1.354
$sv_t \vert s_t = 2$	0.079	0.079	0.127	-0.169	0.333
		dp_{t+1}			
$const_t \vert s_t = 1$	-0.154	-0.154	0.055	-0.262	-0.048
$const_t \vert s_t = 2$	-0.050	-0.050	0.024	-0.096	-0.003
$corp_t \vert s_t = 1$	-0.004	-0.004	0.001	-0.007	-0.001
$corp_t \vert s_t = 2$	-0.001	-0.001	0.001	-0.003	0.000
$dp_t \vert s_t = 1$	0.975	0.975	0.011	0.953	0.997
$dp_t \vert s_t = 2$	0.995	0.995	0.004	0.987	1.004
$sv_t \vert s_t = 1$	-0.013	-0.013	0.005	-0.024	-0.003
$sv_t \vert s_t = 2$	-0.004	-0.004	0.002	-0.009	0.000
		sv_{t+1}			
$const_t \vert s_t = 1$	-0.252	-0.253	0.273	-0.791	0.283
$const_t \vert s_t = 2$	-0.016	-0.016	0.041	-0.098	0.064
$corp_t \vert s_t = 1$	-0.006	-0.006	0.006	-0.018	0.006
$corp_t \vert s_t = 2$	-0.001	-0.001	0.001	-0.003	0.002
$dp_t \vert s_t = 1$	0.012	0.011	0.053	-0.092	0.116
$dp_t \vert s_t = 2$	-0.002	-0.002	0.007	-0.015	0.012
$sv_t \vert s_t = 1$	0.951	0.951	0.025	0.904	1.002
$sv_t \vert s_t = 2$	0.999	0.999	0.004	0.991	1.007

Table 8: This table shows the posterior distribution of parameters for the MS-VAR-TVTP model. From this posterior distribution we present the mean, median, standard deviation and the HPD estimates of the 5th and 95th percentiles. Parameters are sampled on a monthly basis with the sampling period starting from July-1952 until December 2013.

MS-VAR-TVTP Coefficient Posterior Distribution

Coeff.	Post. mean	Post. median	Post. std	5th perc.	95th perc.
$Var(r_t \mid s_t = 1)$	31.882	31.551	3.955	24.723	39.929
$Var(r_t \mid s_t = 2)$	12.703	12.672	0.876	10.977	14.428
$Cov(corp_t, r_t \mid s_t = 1)$	4.102	4.013	1.852	0.469	7.727
$Cov(corp_t, r_t \mid s_t = 2)$	1.927	1.920	0.335	1.265	2.571
$Var(corp_t \mid s_t = 1)$	14.121	13.945	1.861	10.664	17.796
$Var(corp_t \mid s_t = 2)$	3.753	3.742	0.267	3.227	4.279
$Cov(dp_t, r_t \mid s_t = 1)$	-0.314	-0.311	0.039	-0.391	-0.241
$Cov(dp_t, r_t \mid s_t = 2)$	-0.126	-0.126	0.009	-0.143	-0.109
$Cov(dp_t, corp_t \mid s_t = 1)$	-0.039	-0.038	0.018	-0.075	-0.003
$Cov(dp_t, corp_t \mid s_t = 2)$	-0.019	-0.019	0.003	-0.026	-0.013
$Var(dp_t \mid s_t = 1)$	0.003	0.003	0.000	0.002	0.004
$Var(dp_t \mid s_t = 2)$	0.001	0.001	0.000	0.001	0.001
$Cov(sv_t, r_t \mid s_t = 1)$	-0.556	-0.548	0.131	-0.815	-0.305
$Cov(sv_t, r_t \mid s_t = 2)$	-0.021	-0.021	0.011	-0.042	0.000
$Cov(sv_t, corp_t \mid s_t = 1)$	0.075	0.074	0.078	-0.076	0.228
$Cov(sv_t, corp_t \mid s_t = 2)$	-0.007	-0.007	0.005	-0.018	0.003
$Cov(sv_t, dp_t \mid s_t = 1)$	0.005	0.005	0.001	0.003	0.008
$Cov(sv_t, dp_t \mid s_t = 2)$	0.000	0.000	0.000	0.000	0.000
$Var(sv_t \mid s_t = 1)$	0.064	0.063	0.009	0.049	0.081
$Var(sv_t \mid s_t = 2)$	0.004	0.004	0.000	0.003	0.004

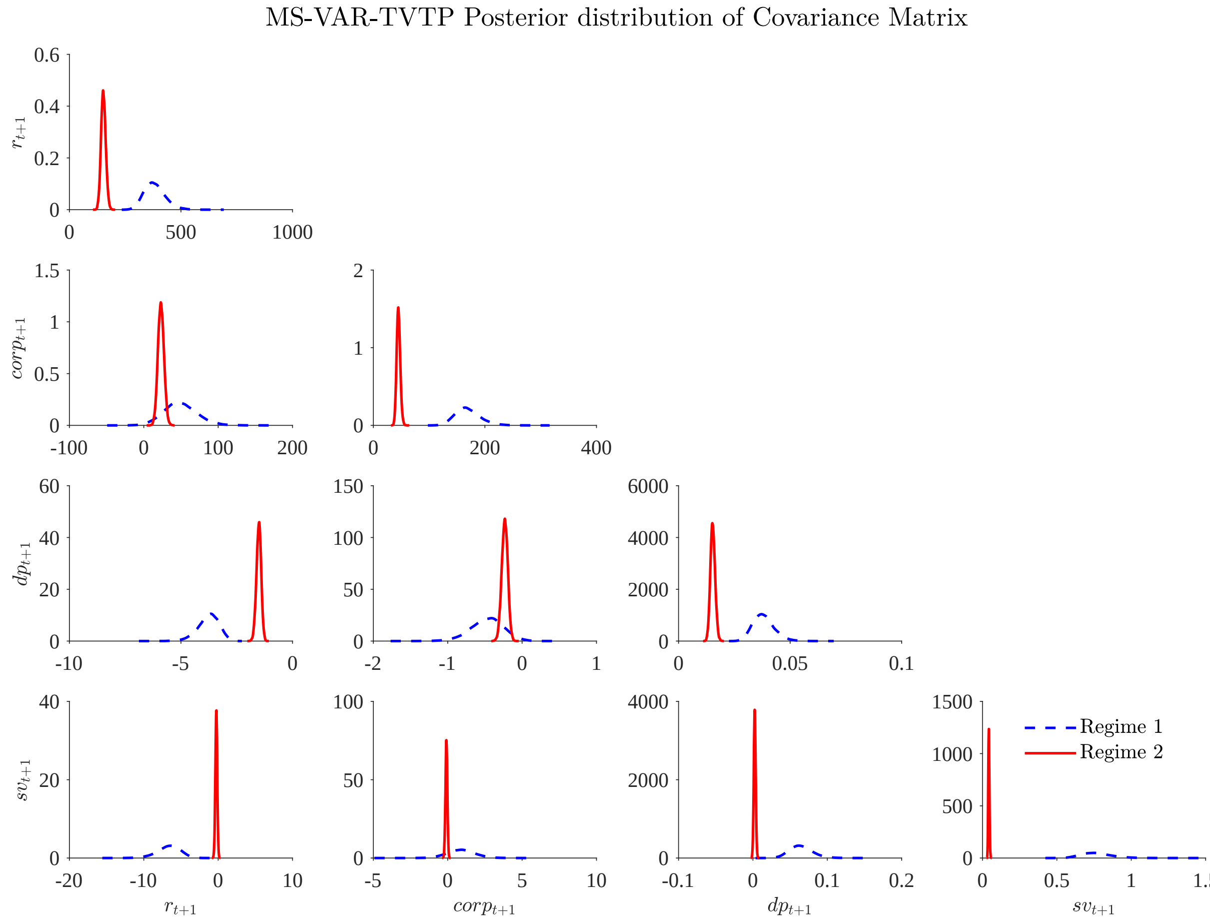

Figure 13: This figure shows the posterior distribution of covariances for the MS-TVTP model conditioned on both regimes. The posterior density plots are shown for the lower triangular covariance matrix with labels on Y-Axis and X-Axis denoting the covariance between response variables, the diagonal elements are the variance density plots of each response variable.

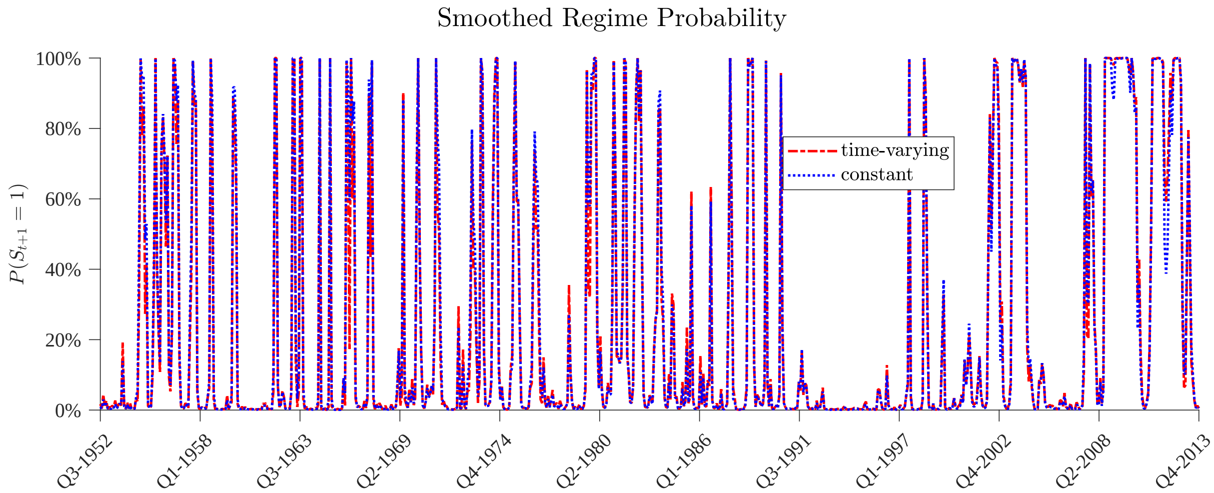

Figure 14: This figure shows the smoothed state probability for $S_t = 1$ for both models. The Y-axis denotes the smoothed probability and the X-axis runs from July-1952 until December 2013.