Abstract

We analyze the effect of model instability on long-term investors using a time-varying VAR(1) model that we estimate using Bayesian Markov Chain Monte Carlo techniques for state-space specifications. Our model is able to handle time-varying intercepts, time-varying slopes, time-varying volatility, time-varying correlation, the leverage effect and fat tails. We calculate the optimal portfolios of long-term investors using numerical Monte Carlo techniques. Time-variation in intercepts and slope coefficients is not persistent enough to be important for long-term investors, while time-variation in the error covariance matrix (especially error volatility) is persistent and therefore very important for long-term investors. Fat tails disappear once time-varying volatility is incorporated. Random walk specifications (persistence equal to 1) or regime-switching models (same persistence for all parameters) lead to a large overestimation of perceived stock market risk and an underinvestment in the stock market. Results are robust to changes in the specification.

Keywords: Strategic asset allocation, time-varying parameters, stochastic volatility, state space model, Bayesian MCMC techniques

JEL: G11, C11, C32, C58, C63
1 Introduction

Long-term investors face substantial uncertainty when modeling future asset returns. Firstly, investors need to select a model to model the dynamics of asset returns. This model can be wrong and investors therefore face model uncertainty. Avramov (2002) shows that this effect is important for both short and long-term investors. Secondly, upon choosing a model long-term investors need to estimate the model parameters. They face parameter uncertainty, since the true parameters are unknown. Parameter uncertainty is very important at long horizons as argued in Barberis (2000). Thirdly, even if the true model and true parameters are known, stock returns are still uncertain due to unexpected shocks (error term). This component is the most important component at short horizons.

Recently, there has been a lot of interest in the strategic asset allocation literature due to the finding that asset returns, specifically stock returns, might be predictable.\textsuperscript{1} If asset returns are predictable, the optimal asset allocations of long-term investors deviate from the allocations of short-term investors. Empirically, it is found by e.g. Campbell, Chan, and Viceira (2003) that long-term investors should invest more in stocks than short-term investors, since stocks mean-revert and are therefore safer in the long-run. This conclusion is robust to the inclusion of parameter uncertainty as shown in Barberis (2000).

The common practice in the strategic asset allocation literature is to estimate a model on a data-set of 50 years or more and to assume that model coefficients remain constant over this period. However, a priori there is no reason to believe that they indeed are. Aspects such as institutional changes, wars or changes in the stock market behavior of participants due to different risk aversion levels or different financial sophistication levels could lead to changes in the relation between asset returns and predictor variables or to changes in the properties of the error term.

While small changes in coefficients might not have a large impact on short-term investors, they could have a large impact on long-term investors if they are long-lasting and persistent. In that case, the mistakes one makes in using a constant model add up over the investment horizon and can become very large. If on the contrary the changes in coefficients are not of persistent but of transitory nature, they are unlikely to be important for long-term investors. In that case the coefficients only deviate from the constant model for a fraction of the investment horizon. Hence, the persistence of time-varying parameters is extremely important at long horizons.

Another common practice in the strategic asset allocation literature is to impose Gaussian error distributions, whereas evidence clearly shows that the error distributions deviate significantly from normality. Risk-averse investors evaluate very good and very bad outcomes differently, since they want to avoid bad outcomes at all costs. They value models that are able to avoid these bad outcomes. This suggests that the normality assumption could lead to very misguided investment advice, since it does not accurately model extreme tail events. Properly modeling the error term is therefore very important for risk-averse investors.

\textsuperscript{1}Campbell and Viceira (2002) and Brandt (2010) give an overview of the strategic asset allocation literature.
We consider the effect of model instability on long-term investors using a time-varying VAR(1) model in which parameters are allowed to change in every period. We develop a methodology that is able to handle time-varying intercepts, time-varying slope coefficients, time-varying error volatility, time-varying error correlation, the leverage effect and fat tails. We assess the importance of these model components for long-term investors. We focus on the persistence of the different time-varying parameters.

An alternative specification is the regime-switching model that Guidolin and Timmermann (2007) and Pettenuzzo and Timmermann (2010) implement. We do not pursue this alternative here because of three reasons. Firstly, a priori it seems more likely that the behavior of stock market participants changes smoothly over time (due to changes in risk aversion or financial sophistication) instead of abruptly, which suggests that a regime-switching model is not appropriate. Secondly, one of our objectives is to assess what kind of model instability is the most important for long-term investors. A regime-switching model does not allow us to assess the individual components, since all components change jointly. Finally, a regime-switching model pools the persistence parameter of all components. In other words, changes in say the slope of stock returns are equally persistent as changes in the volatility of a predictor variable, while there is a priori no reason to impose such a restriction.

A related paper is Johannes, Korteweg, and Polson (2011). These authors develop particle filtering techniques to assess how the views of economic decision makers evolve over time using a VAR model in which only some parameters are allowed to change over time. Our paper differs in both scope and perspective. Firstly, we use smoothed estimates to assess ex-post whether there is model instability, while Johannes, Korteweg, and Polson (2011) use filtered estimates. Since we use smoothed parameters, we are able to assess more efficiently whether there was time variation in the past 82 years. Since Johannes, Korteweg, and Polson (2011) use filtered estimates, they would find time-variation even if there was none. Secondly, we consider time-variation in all intercepts and slope coefficients, add the leverage effect, consider non-normal distributions and also consider time-varying error correlation. Johannes, Korteweg, and Polson (2011) only consider time-variation in one slope coefficient and in the volatility of the error term. Thirdly, we look at specifications that also consider bond returns and are therefore more relevant for long-term investors and indeed find interesting effects in the correlation between stocks and bonds.

We find that it is important for long-term investors to take model instability into account. CER gains are up to 5% per year. Long-term investors should take time-varying volatility and correlation into account, but can safely ignore time-varying slopes and excess kurtosis once time-varying volatility is incorporated. The reason is that time-varying slopes are not persistent enough to be of importance for long-term investors and that fat tails are not important once stochastic volatility is incorporated. The persistence of time-varying parameters is extremely important and a random walk specification (persistence equal to 1) or a regime-switching model (same persistence for all parameters) is therefore not appropriate. Results are robust to changes in the main specification.
This paper is organized as follows. Section 2 discusses the data-set we use and performs a preliminary analysis. Next, section 3 explains the methodology. It discusses the model, the Bayesian prior distribution and the Bayesian MCMC techniques. Section 4 shows the results for the basic specification in which the dividend-to-price ratio is incorporated as predictor of asset returns. Section 5 performs a robustness check using the yield-spread as predictor variable. Finally, section 6 provides the conclusion. The appendix contains details on the numerical techniques we use to estimate the model.

2 Data and preliminary analysis

We use a monthly data-set that starts in December 1926 and ends in December 2008 for the US stock and bond markets. It is based on Goyal and Welch (2008). We use three asset returns and two predictors.

The first asset return is the ex post real T-bill rate \((R_{\text{tbill}})\) which we obtain by subtracting log inflation from the log return on the 3-month T-bill rate. We do not include the T-bill rate in the econometric models to keep our models parsimonious, but we use its average value, which is 0.060\% per month, in portfolio construction. The second asset return is the excess log stock return \((X_s)\). It is defined as the difference between the log return (including dividends) on the S&P 500 and the log return on the (nominal) 3 month T-bill. The third asset return is the excess log return on a nominal long-term government bond (maturity of approximately 20 years) and is defined in a similar way.

Our data-set contains two predictor variables. The first predictor variable is the log dividend-to-price ratio, defined as the log difference between dividends over the past four quarters and the current S&P index level. Campbell and Shiller (1998) and Cochrane (2007) show that this ratio is an important predictor of stock returns. Secondly, we consider the yield-spread, which is the difference between the log yield on a long-term government bond and the log yield on the 90-day T-bill. It is an important predictor of both stock and bond returns, refer to Campbell (1995) and Fama and French (1989).

Table 1 shows the summary statistics for the data-set. Firstly, the equity risk premium of 5.4\% per year is in line with other papers. Secondly, the kurtosis and skewness clearly indicate that the variables - especially stock and bond returns - deviate significantly from normality. Thirdly, the AR(1) coefficients indicate that the predictor variables are very persistent.

We consider two different specifications. The first specification is a VAR(1) model in which stock returns, bond returns and the dividend yield are regressed on a constant and the lagged dividend yield. This specification is considered in section 4. The second specification is a VAR(1) model in which excess stock returns, bond returns and the yield spread are regressed...
on a constant and the lagged yield spread. We consider this model in the robustness section 5. We do not consider VAR(1) models of dimension greater than 3 to keep our analysis feasible and to reduce the total number of parameters.

Table 2 shows the OLS estimates, standard errors, the covariance matrix of the residuals and the skewness and kurtosis of the residuals for both specifications. We find that the dividend-to-price ratio (panel A) and the yield spread (panel B) are positively related to excess stock returns. These positive coefficients combined with the negative error correlation between excess stock returns and especially the dividend-to-price ratio suggest that stocks mean-revert on average. The large standard errors indicate that there is a lot of estimation uncertainty involved. Next, the yield spread is a positive and strong predictor of excess bond returns, but hardly predicts excess stock returns. Its positive coefficient combined with the negative error correlation between excess bond returns and the yield spread suggests that bond returns show some mean-reversion. Finally, the skewness and kurtosis values indicate that the error terms deviate strongly from normality. The kurtosis value for the yield-spread is especially remarkable and needs to be further analyzed.

3 Methodology

In this paper, we use a first order time-varying Vector Autoregression - TVAR(1) - to model the investment opportunity set of long-term investors.\(^2\) The model is able to handle time-varying intercepts and slope coefficients, a time-varying error covariance matrix with both volatility and correlation time-varying, a leverage effect in volatility and finally error terms with fat tails. The model is estimated by Bayesian MCMC techniques.\(^3\)

Our methodology is an extension of the methodology of Primiceri (2005). We extend Primiceri’s (2005) methodology by allowing for fat tailed error distributions, by estimating the leverage effect in volatility and by estimating AR(1) processes for all transition equations. These extensions are very relevant given the data-set we use. Firstly, it is well-known (e.g. Omori, Chib, Shephard, and Nakajima (2007), Jacquier, Polson, and Rossi (2004)) that the leverage effect in volatility and fat tailed error distributions are present in data on stock returns. Secondly, empirically we find that it is very important to estimate the persistence of time-varying processes instead of imposing random walks. We explain below that random walks lead to a large overestimation of perceived risk of long-term stock returns. Ignoring these extensions would lead to a misspecified model.

\(^2\)Time-variation in the model might be a sign of time-variation in the parameters of the DGP or a sign of misspecification of the model. Even if it is a sign of misspecification, it does not invalidate the use of the time-varying model. All models are wrong, but some are actually useful. The TVAR(1) is a very flexible and therefore useful model.

\(^3\)In theory, it would be possible to estimate the model with frequentist techniques. However, this would be extremely difficult due to the large number of parameters and the non-linearity of the problem. Bayesian methods on the other hand are well-suited to estimate a problem of this magnitude, since it divides the original estimation problem in smaller and simpler steps (the Gibbs step).
3.1 Model

In this section, we explain the model. Define the \( n \times 1 \) vector \( y_t \) as follows (\( n = 3 \) in the empirical application)

\[
y_t = \begin{pmatrix} x_t \\ z_t \end{pmatrix},
\]

(1)

where \( x_t \) is a \( n - k \times 1 \) vector consisting of the asset returns and \( z_t \) a \( k \times 1 \) vector of predictor variables at time \( t \). The model is as follows

\[
y_t = a_t + B_t z_{t-1} + u_t,
\]

(2)

for \( t = 1, ..., T \), where \( a_t \) is a \( n \times 1 \) vector of intercepts, \( B_t \) an \( n \times k \) matrix of slope coefficients and \( u_t \) an \( n \times 1 \) vector of error terms with covariance matrix \( \Omega_t \) whose properties are shown below.

First, we introduce some additional notation. Let \( X_t = I_n \otimes [1, z_{t-1}] \) and let \( b_t = vec ([a_t, B_t]) \). The model can be rewritten as

\[
y_t = X_t b_t + u_t.
\]

(3)

Without loss of generality, we consider a triangular reduction of covariance matrix \( \Omega_t \)

\[
L_t \Omega_t L_t' = \Sigma_t \Sigma_t,
\]

(4)

where \( L_t \) is an \( n \times n \) lower triangular matrix

\[
L_t = \begin{pmatrix} 1 & 0 & 0 & 0 \\ l_{21,t} & 1 & 0 & 0 \\ \vdots & \ddots & 1 & 0 \\ l_{n1,t} & \ldots & l_{n-1,t} & 1 \end{pmatrix}
\]

(5)

and where \( \Sigma_t \) is an \( n \times n \) diagonal matrix

\[
\Sigma_t = \begin{pmatrix} \sigma_{t,1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_{t,n} \end{pmatrix}.
\]

(6)

The common assumption in the literature (e.g. Campbell, Chan, and Viceira (2003), Barberis (2000)) is to assume that the error term is normally distributed. In that case, we could write

\[
u_t = L_t^{-1} D_t e_t,
\]

where \( e_t \) has a standard normal distribution. Since \( \Omega_t = L_t^{-1} \Sigma_t L_t L_t'^{-1} \), \( u_t \) would be distributed

\[
\text{normal}(0, \Sigma_t),
\]

Due to the triangular reduction, the results depend in theory on the ordering of the variables. Our empirical results turn out to be robust to different orderings.
as $N(0, \Omega_t)$.

However, there is ample evidence that suggests that the distribution of the error term deviates from normality, see for example the preliminary results in the previous section. Therefore, we consider an alternative specification for the error term and assume that $u_t$ has a distribution with fat tails. Therefore, we introduce scale mixture variables $\lambda_{t,1}, \ldots, \lambda_{t,n}$ and $n \times n$ diagonal matrix

$$\Lambda_t = \begin{pmatrix}
\lambda_{t,1}^\frac{1}{2} & 0 & 0 \\
0 & \ddots & 0 \\
0 & 0 & \lambda_{t,n}^\frac{1}{2}
\end{pmatrix}$$

and make the following assumption for $u_t$

$$u_t = L_t^{-1} \Sigma_t \lambda_t e_t,$$

where the elements of $e_t$ have independent standard normal distributions.

By multiplying $e_{t,j}$ by scale mixtures $\lambda_{t,j}^\frac{1}{2}$, error term $u_{t,j}$ deviates from normality $\forall j$. We follow Omori, Chib, Shephard, and Nakajima (2007) and use the following distribution for $\ln \lambda_{t,j}$ for $j = 1, \ldots, n$

$$\ln \lambda_{t,j} \sim N\left(-\frac{1}{2} \tau_j^2, \tau_j^2\right).$$

If $\tau_j^2 = 0$, $\lambda_{t,j}^{\frac{1}{2}} e_{t,j}$ would have a normal distribution. If $\tau_j^2 > 0$, $\lambda_{t,j}^{\frac{1}{2}} e_{t,j}$ has a normal log-normal distribution.\footnote{An alternative would be to specify that $\lambda_{t,j}$ has an inverse gamma $i\Gamma(a, \nu)$ distribution such that $\lambda_{t,j}^{\frac{1}{2}} e_{t,j}$ would have a student-t distribution with $\nu$ degrees of freedom. However, this choice leads to an unstable numerical algorithm and therefore we do not pursue this alternative in this paper.} This is a distribution with fat tails. Its moments for different values of $\tau_j^2$ are given in table 3. The table clearly shows that the multiplication of $e_{t,j}$ for $j = 1, \ldots, n$ with the scale mixture only impacts the kurtosis. The mean, standard deviation and skewness are not influenced. Therefore, $\Omega_t$ can still be interpreted as the covariance matrix of $u_t$.

[Table 3 about here.]

Next, we specify the dynamics of the time-varying parameters. Let $l_t$ be the $n(n - 1)/2 \times 1$ vector of non-zero and non-one elements of $L_t$ (stacked by rows) and let $\sigma_t$ be the $n \times 1$ vector of diagonal elements of $\Sigma_t$. The evolution of the time-varying parameters in the model - $b_t$, $l_t$ and $\sigma_t$ - is modeled as follows

$$b_{t+1} - \mu_b = A_b (b_t - \mu_b) + \eta_t$$

$$l_{t+1} - \mu_l = A_l (l_t - \mu_l) + \zeta_t$$

$$\ln \sigma_{t+1}^2 - \mu_s = A_s (\ln \sigma_t^2 - \mu_s) + \xi_t,$$

where $\mu_b$, $\mu_l$ and $\mu_s$ are respectively the unconditional means of $b_t$, $l_t$ and $\ln \sigma_t^2$, where $A_b$, $A_l$ and $A_s$ are the transition matrices and where $\eta_t$, $\zeta_t$ and $\xi_t$ are the innovations. The innovations
are distributed as $\eta_t \sim N(0, Q)$, $\zeta_t \sim N(0, R)$ and $\xi_t \sim N(0, S)$. The next subsection explains that we need to impose restrictions on $R$ and $S$ to make the analysis tractable.\footnote{In order to estimate the model, we use the simulation smoother of Durbin and Koopman (2002). We use their timing convention for the innovations in the transition equations.}

We follow Primiceri (2005) and assume that $\eta_t$, $\zeta_t$ and $\xi_t$ are independent of each other and that $\eta_t$ and $\zeta_t$ are independent of error term $e_t$. However, unlike Primiceri (2005) we do allow for a leverage effect in volatility by specifying that the correlation between $e_{t,j}$ and $\xi_{t,j}$ is (instead of 0)

$$\text{corr}(e_{t,j}, \xi_{t,j}) = \rho_j, \forall j.$$  \hspace{1cm} (11)

Omori, Chib, Shephard, and Nakajima (2007) and Jacquier, Polson, and Rossi (2004) find that this correlation is negative for a shock to stock returns and an innovation in its stochastic volatility. This means that if there is a negative shock to stock returns at time $t$, next period’s log volatility $\ln \sigma_{t+1}^2$ will be higher on average.

We introduce some additional notation. Firstly, define $\tilde{b}_{t+1} = b_{t+1} - \mu_b$, $\tilde{t}_{t+1} = t_{t+1} - \mu_t$ and $\tilde{\ln} \sigma_{t+1}^2 = \ln \sigma_{t+1}^2 - \mu_s$. Secondly, let $b$, $l$ and $\ln \sigma^2$ be vectors that stack all values of $b_t$, $l_t$ and $\ln \sigma_t^2$ \forall t. Finally, define $\tau^2$ as the vector that vertically stacks the values for $\tau^2_j \forall j$ and define vector $\rho$ similarly for $\rho_j \forall j$.

It is very important to note that say $e_{t,3}$ is not the error term for the third equation of the system in period $t$. Instead, it is the part of the third equation error term that is orthogonal to the errors of the first two equations. Likewise, $\tau^2_3$ and $\rho_3$ are not the excess kurtosis and correlation coefficients for the third equation, but are the excess kurtosis and correlation coefficients for the part of the third equation error term that is orthogonal to the first two equations.

We consider several alternative models. Firstly, we impose some or all of the following restrictions

- Time-constant $b_t : Q = 0$
- Time-constant $l_t$ and $\ln \sigma_{t+1}^2: R = 0, S = 0$
- Normally distributed error terms: $\tau^2 = 0$

Secondly, we consider a pooled model in which the persistence of all time-varying parameters is equal

$$a_{b(i,i)} = \rho_p, \forall i = 1, ... nK$$
$$a_{l(j,j)} = \rho_p, \forall j = 1, ... n(n - 1)/2$$
$$a_{s(k,k)} = \rho_p, \forall k = 1, ..., n$$

and 0 otherwise.

Finally, we impose random walks (without drift) by setting the persistence of all time-varying
parameters equal to 1
\[ b_{t+1} = b_t + \eta_t \]
\[ l_{t+1} = l_t + \zeta_t \]
\[ \ln \sigma_{t+1}^2 = \ln \sigma_t^2 + \xi_t. \]

Note that in such models, the unconditional distribution of time-varying parameters does not exist.

These alternative models are either nested in our most general model or are (in the case of the random walk specification) straightforward extensions. Since they do not lead to any further issues regarding prior choice, posterior distribution or MCMC algorithm, we do not explicitly deal with them in the next two subsections.

3.2 Prior

In this section, we explain the prior distributions for the most general model. It is an hierarchical model and therefore we have to define prior distributions for the initial conditions \( \tilde{b}_1, \tilde{l}_1 \) and \( \tilde{\ln} \sigma_1^2 \) and for the hyperparameters of the model (\( \mu_b, \mu_l, \mu_s, A_b, A_l, A_s, Q, R, S, \rho, \tau^2 \)). We use the first 60 months of our data-set as a training sample to estimate a time-constant VAR(1) model by OLS and use its estimates in the construction of some of the prior distributions below.

Firstly, we assume that the initial conditions are drawn from their stationary unconditional distribution
\[ p(\tilde{b}_1 | A_b, Q) = N(0, \Sigma_b) \]  
\[ p(\tilde{l}_1 | A_l, R) = N(0, \Sigma_l) \]  
\[ p(\tilde{\ln} \sigma_1^2 | A_s, S) = N(0, \Sigma_s), \]

where \( \text{vec}(\Sigma_b) = (I_{(nK)^2} - A_b \otimes A_b)^{-1} Q \) and where \( \Sigma_l \) and \( \Sigma_s \) are defined similarly.

Secondly, we consider the transition matrices. We set the off-diagonal elements of the matrices equal to 0 and assume the following for the diagonal elements
\[ p(A_{b,(i,i)}) = N(m_b, v_b)I(A_{b,(i,i)}), \forall i = 1, ..., nK \]  
\[ p(A_{l,(j,j)}) = N(m_l, v_l)I(A_{l,(j,j)}), \forall j = 1, ..., n(n - 1)/2 \]  
\[ p(A_{s,(k,k)}) = N(m_s, v_s)I(A_{s,(k,k)}), \forall k = 1, ..., n, \]

where
\[ m_b = m_l = m_s = 0.9 \]
\[ v_b = v_l = v_s = 0.2^2 \]

and where \( I(x) \) is equal to 1 if \(-1 < x < 1\) and 0 otherwise. Hence, we impose stationary
processes for the time-varying parameters. The prior correlations between the elements are equal to 0.

Thirdly, we use the training sample to specify the prior distributions for the unconditional means as follows\(^7\)

\[
p(\mu_b) = N(\hat{b}_{OLS}, 10^6 V(\hat{b}_{OLS}))
\]
\[
p(\mu_l) = N(\hat{l}_{OLS}, 10^6 V(\hat{l}_{OLS}))
\]
\[
p(\mu_s) = N(\ln \hat{\sigma}^2_{OLS}, 10^6 V(\ln \hat{\sigma}^2_{OLS})).
\]

Mean \(\hat{b}_{OLS}\) and covariance matrix \(V(\hat{b}_{OLS})\) are respectively the standard OLS estimates and its covariance matrix. We do not have direct OLS estimates for \(\mu_l\) and \(\mu_s\) and its covariance matrices, but instead have an estimate of the covariance matrix of the residuals \(\hat{E}\). We draw this covariance matrix 1000 times from the inverse Wishart distribution \(\text{iWishart}(\hat{E}' \hat{E}, T)\), construct \(l_{OLS}\) and \(\ln \sigma^2_{OLS}\) for every draw and use its means and covariance matrix across the 1000 draws to calculate \(\hat{l}_{OLS}, V(\hat{l}_{OLS}), \ln \hat{\sigma}^2_{OLS}, V(\ln \hat{\sigma}^2_{OLS})\).

It is common in the literature (e.g. Primiceri (2005)) to use the variance of the estimated intercepts, slopes, volatilities and \(\hat{l}_{OLS}\)’s of the time-constant model to set-up the covariance matrices of the innovations of the transition equations. We follow this trend and set the mean of the inverse Wishart and inverse gamma distributions equal to a constant fraction \(c\) of the covariance matrices of the estimated coefficients. For the degrees of freedom, we choose the minimum degrees of freedom such that the prior means actually exist. Constant \(c\) is specified below.

The prior distribution for \(Q\) is

\[
p(Q) = \text{iWishart}(W_Q, df_Q),
\]

where

\[
W_Q = c \times (df_Q - nK - 1) \times D_Q
\]
\[
df_Q = nK + 2
\]

where \(D_Q\) is a diagonal matrix with the diagonal elements of \(V(\hat{b}_{OLS})\) on the diagonal. It is common in the literature to impose a diagonal matrix for \(Q\) itself to reduce the total number of parameters, while there is a priori no reason to expect that the off-diagonal elements are equal to 0. In our setting, \(Q\) can be any positive definite matrix, but is shrunk towards a diagonal matrix by choosing this particular prior scale matrix. In this way, we try to find the balance between flexibility and efficiency.

There is no good guidance on how to choose the multiplication constant \(c\). If we choose

\(^7\)Although the parameters are unlikely to be exactly constant in the training sample, we assume that time-variation is limited over such a short time-span. Furthermore, we choose the priors as uninformative as possible such that they only have a negligible impact on the posterior distribution.
c too large, our prior implies too much time-variation. If we choose c too low, the simulation algorithm does not work smoothly. We choose c equal to 0.01. Results are not noticeably different to specifications with c = 0.1 or c = 0.001. In this way, the prior is not flat, but still diffuse and relatively uninformative. This same constant c is chosen in the priors for R and S.

In order to obtain partially analytical results for $l_t$, we need to impose a blockdiagonal structure for matrix $R$ as in Primiceri (2005). Since $n = 3$, the matrix contains two blocks. The first block is formed by element $R_{1,1}$, i.e. the variance of innovations to $l_{t,(1)}$. The second block is formed by elements $R_{2:3,2:3}$, i.e. the covariance matrix of innovations to $l_{t,(3,1)}$ and $l_{t,(3,2)}$. This gives the following prior distributions

$$p(R_{(1,1)}) = i\text{Gamma}(W_{r1}, df_{r1})$$

$$p(R_{(2:3,2:3)}) = i\text{Wishart}(W_{r2}, df_{r2}),$$

where

$$W_{r1} = c \times (df_{r1} - 2) \times D_{l,1}$$

$$W_{r2} = c \times (df_{r2} - 3) \times D_{l,2}$$

$$df_{r1} = 1 + 2$$

$$df_{r2} = 2 + 2$$

and where $D_{l,1}$ is $V(\hat{l}_{OLS,(1,1)})$ and $D_{l,2}$ is a diagonal matrix with the diagonal elements of $V(\hat{l}_{OLS,(2:3,2:3)})$ on its diagonal. As above, matrix $R_{(2:3,2:3)}$ can be any positive definite matrix, but we shrink it towards a diagonal matrix to get more efficient estimates.

Next, let us consider the prior covariance matrix for $S$ and the prior for $\rho$ jointly. In order to make the analysis tractable, we choose a diagonal matrix for $S$. This allows us to consider the three stochastic volatility equations separately. Let us consider the covariance matrix $\Sigma^*_j$ of $e_{t,j}$ and $\xi_{t,j}$ jointly for $j = 1, \ldots, n$

$$\Sigma^*_j = \left( \begin{array}{cc}
\frac{1}{\sqrt{\rho_j S_{j,j}}} & \rho_j \sqrt{S_{j,j}} \\
\rho_j \sqrt{S_{j,j}} & \frac{S_{j,j}}{S_{j,j}}
\end{array} \right).$$

It is difficult to formulate a prior for $\Sigma^*_j$, since its (1,1) element is equal to 1. Therefore, we reparameterize $S_{j,j}$ and $\rho_j$ to be able to choose prior distributions in the way proposed in Jacquier, Polson, and Rossi (2004)

$$\Sigma^*_j = \left( \begin{array}{cc}
\frac{1}{\psi_j} & \psi_j \\
\psi_j & \theta_j + \psi_j^2
\end{array} \right).$$

As in Jacquier, Polson, and Rossi (2004), we choose an inverse gamma prior for $\theta_j$ and a normal
prior for $\psi_j|\theta_j$ such that we obtain a tractable algorithm. We get for $j = 1, \ldots, n$

\begin{align*}
    p(\theta_j) &= i\text{Gamma}(W_{\theta,j}, df_\theta) \\
    p(\psi_j|\theta_j) &= N(0, \theta_j/p),
\end{align*}

where

\begin{align*}
    W_{\theta,j} &= c(df_\theta - 2)V(\ln \sigma^2_{OLS,(j,j)}) \\
    df_\theta &= 3 \\
    p &= 3
\end{align*}

Clearly, $S_{j,j} = \theta_j + \psi_j^2$ and $\rho_j = \frac{\psi_j}{\sqrt{\theta_j + \psi_j^2}}$ for $j = 1, \ldots, n$.

Finally, we choose a prior distribution for $\tau^2_j$ for $j = 1, \ldots n$. As in Omori, Chib, Shephard, and Nakajima (2007) we use a gamma prior

\begin{equation}
    p(\tau^2_j) = \text{Gamma}(a_\tau, b_\tau)
\end{equation}

with

\begin{align*}
    a_\tau &= 0.25 \\
    b_\tau &= 2
\end{align*}

This prior has a mean of 0.5 and a relatively large variance of 1.

### 3.3 Posterior and MCMC

In this section, we sketch how we estimate the general model. Exact details are given in the appendix. The simulation algorithm is more complicated than Primiceri (2005) due to the formulation of the initial conditions in equations (14), (15) and (16), the estimation of transition matrices, the estimation of the leverage effect and the presence of fat-tailed error distributions.

We use Markov Chain Monte Carlo techniques to obtain posterior distributions of the parameters of interest. In general, time-varying parameters and their unconditional means are drawn using the Kalman filter - smoother technique of Durbin and Koopman (2002). Transition matrices and covariance matrices are simulated using Metropolis-Hastings steps.

The system for $b$ is a linear Gaussian state space model (conditional on $l$, $\sigma$ and the hyperparameters) and $b$ and $\mu_b$ can therefore be easily simulated. The system for $l$ is in general not a linear Gaussian state space model (conditional on $b$, $\sigma$ and hyperparameters), but can be transformed to $n - 1$ linear Gaussian state space models if we impose that $R$ is blockdiagonal. Under this assumption, sampling $l$ and $\mu_l$ is straightforward. The system for $\sigma^2$ is not a linear Gaussian state space model (conditional on $b$, $l$ and hyperparameters), but can be transformed into $n$ approximately linear Gaussian state space models by transforming $\sigma^2_t$ to $\ln \sigma^2_t$ and by
approximating the errors of the observation equation of the state space model by a mixture of normals as suggested in Kim, Shephard, and Chib (1998) and Omori, Chib, Shephard, and Nakajima (2007). In this way, we can still use the simulation smoother of Durbin and Koopman (2002) to simulate $\ln \sigma^2$ and $\mu_s$.

If the initial conditions in equations (14), (15) and (16) would not depend on the transition matrices and covariance matrices, we could easily simulate them using a Gibbs sampler by respectively a normal distribution and an inverse Wishart distribution. Since they do, we use these distributions as proposal densities in a Metropolis-Hastings step. Since the acceptance probabilities are all larger than 80%, this hardly deteriorates the sampling performance of the MCMC algorithm.

We do not sample the correlations $\rho$ directly, but instead consider the transformation in equation (30) to make the sampling step easier. These transformed parameters are sampled using a Metropolis-Hastings algorithm using inverse gamma and normal proposal densities.

Finally, the conditional posterior distribution for $\tau_j^2$, for $j = 1, ..., n$ is not a known distribution. Therefore, we use another Metropolis-Hastings step. We sample $\ln \tau_j^2$, because its posterior distribution is easier to approximate by a student-t proposal density. This gives an acceptance probability of 97%.

In the empirical section, we retain 10,000 iterations after a burn-in period of 5,000 iterations. We draw 10 asset return paths per iteration (hence 100,000 in total) to calculate predictive distributions and portfolio weights. Increasing the number of iterations does not significantly impact results.

4 Results basic specifications

We report results for the basic specification with the dividend-to-price ratio as predictor. We explain estimation results, consider the term structures of risk and portfolio weights for the time-varying model, provide an assessment of the individual components of the time-varying model and finally analyze the importance of the persistence of time-varying parameters.

4.1 Estimation results

Figure 1 plots the time-series of (smoothed) posterior means of the time-varying intercepts and slope coefficients for both the time-varying and constant model. A few results stand out. Firstly, the dividend-to-price ratio is a positive predictor of excess stock and bond returns in every period. Both coefficients are on average quite a bit higher than the ones in the constant model. Secondly, both prediction coefficients show some modest time-variation. For example, the posterior mean in the stock return equation varies between 0.008 and 0.015. The time-variation is however not very persistent, since deviations from means do not last very long. Thirdly, the AR(1) coefficient for the dividend-to-price ratio is on average lower than the one in the constant model and varies over time. Such a change in persistence of a predictor can
potentially have a large impact on long-term investors. The constant in the predictor equation varies a lot over time as well.

[Figure 1 about here.]

[Table 4 about here.]

Table 4 shows the posterior means and standard deviations of the most important hyperparameters of several specifications. The time-varying model is in the last column. The other specifications are considered in subsection 4.3. The first 12 rows show the posterior means and standard deviations of the diagonal elements of $A_{slope}$. The table confirms that time-variation in intercepts and slopes is not very persistent. The average posterior mean of these persistence parameters is around 0.90, which implies a half life of an innovation of only a bit more than 6 months.

We conclude that there is time-variation in intercepts and slopes, but that this time-variation is rather small and not very persistent.

[Figure 2 about here.]

Figure 2 reports the time-series of (smoothed) posterior means of the covariance matrix of the error terms. It reports volatilities on the diagonal, correlations above the diagonal and values for $l_t$ below the diagonal. The figure shows some interesting results. Firstly, the error volatility in the stock return equation varies considerably over time. It ranges from 17% per month in the 1930s to 2.5% in the 1960 and 1970s. The figure also shows that time-variation is very persistent, since volatility reverts slowly to its mean. Secondly, there is a lot of time-variation in the correlation between stocks and bonds. Correlation can be positive as in the 1970s and 1980s, but can also be negative as in more recent times. This time-variation in correlation is also very persistent. The correlation between stock and dividend-to-price ratio innovations also varies somewhat, but stays very close to -1. Thirdly, the error volatility for the dividend-to-price ratio almost exactly mimics the error volatility for stock returns. This is not surprising because of the consistently strong negative correlation between them. Fourthly, the error volatility of bond returns varies a lot over time. It reaches its maximum of 6% per month at the end of the sample. Its error correlation with the dividend-to-price ratio is almost exactly the negative of its error correlation with stocks.

Table 4 shows some interesting results for the error covariance matrix. Firstly, it indicates that time-variation in error volatilities and error correlations are much more persistent than the time-variation in intercepts and slopes. The posterior mean of the persistence parameter of error volatility in the stock return equation is 0.96. This implies a half-life of almost 1 1/2 years. The other error volatilities are even more persistent. Secondly, there is hardly any excess kurtosis left after taking time-variation in volatility into account. For example, parameter $\tau_2^2$ shows that the kurtosis of the error term in the stock equation is very close to 3 instead of more than 10 as
in table 2. Hence, if one ignores time-variation in volatility, one would wrongly conclude that the distribution of error terms deviates substantially from normality.

Posterior mean \( E(\rho_1|Y) = -0.43 \) shows that there is a leverage effect in stock returns at the monthly horizon even if predictors are included. This means that three aspects play a role if there is a negative shock to stock returns. Firstly, a negative innovation leads to higher expected future stock returns due to the mean-reversion effect. Secondly, the negative innovation increases the error volatility for stocks. Thirdly, the negative shock increases the error volatility of the dividend-to-price ratio, since \( e_{t,1} \) is a major component of the error for this ratio.\(^8\)

The first effect is beneficial for long-term investors and implies that long-term investors should invest more in stocks than short term investors as shown in e.g. Campbell, Chan, and Viceira (2003). However, the second and third effect increase the risk for long-term investors and reduce their stock allocations. The preceding suggests that ignoring time-variation in volatility leads to an overinvestment in stocks at long horizons. The second effect is considered in isolation in Chacko and Viceira (2005), who conclude that it leads to a modest negative hedge term for long-term investors. The third effect is ignored in the literature. There are no other papers that consider all three effects jointly as we do.

We conclude that there is a lot of time-variation in error volatility and correlation and that this time-variation is very persistent and relevant for long-term investors.\(^9\)

4.2 Term structure of risk and portfolio weights

Figure 3 reports the term structures of risk for excess log stock returns for both the time-varying and time-constant model. It shows the annualized predictive volatility of future cumulative stock returns. We obtain these figures by simulating stock returns from the predictive distribution of future stock returns. In the upper panel we draw time-varying parameters from its unconditional posterior distribution, draw the time-constant parameters from its posterior distribution and we set the dividend-to-price ratio equal to its historical average. In the bottom panel, we simulate the time-varying parameters from the posterior distribution in December 2008, we simulate the time-constant parameters from its posterior distribution and we set the dividend-to-price ratio equal to its December 2008 value.

The term structures of risk takes parameter uncertainty, state uncertainty and uncertainty due to the error term into account. An investor faces state uncertainty, since she does not

\(^8\)The error for the dividend-to-price ratio in period \( t \) is a linear combination of \( e_{t,1} \), \( e_{t,2} \) and \( e_{t,3} \). The almost perfect negative correlation between the error in the dividend-to-price ratio and the error in stock returns shows that the latter two components hardly matter for the error in the dividend-to-price ratio equation.

\(^9\)A note on methodology. It is well-known in the time-series literature (e.g. Breusch and Pagan (1979)) that one can rewrite an AR(1) model with random coefficients as an AR(1) with heteroscedastic errors. In such a setting, the extra variation due to the random coefficient only leads to a heteroscedastic error term, but does not affect the mean. However, in our setting, the time-varying parameters are autocorrelated over time, i.e. if for example the autocorrelation of stock returns is high today, it will also be (relatively) high tomorrow. Therefore, the time-variation in say the slope parameter in for example an AR(1) does not only affect the conditional variance but also the conditional mean. Therefore, such a model cannot simply be rewritten as an heteroscedastic AR(1).
know the exact values of the time-varying parameters (states) and only knows their posterior distribution at any point in time.

The figure shows that the term structure of risk for the time-varying specification varies a lot over time. It can either be upward sloping (upper panel) or downward sloping (bottom panel). Let us firstly consider the term structure in the bottom panel. The error volatility at the 1-month horizon is very high (27%). Figure 2 already showed that at the end of the sample, the error volatility for stock returns is much higher than its mean. At medium horizons, the annualized predictive volatility is however much lower, because (i) mean-reversion in stock returns makes stocks safer in the long-run and (ii) the error volatility reverts back to its (lower) long-run mean. At horizons of 15 years or more, the term structure is slightly upward sloping due to the effect of parameter uncertainty. At the 20-year horizon, annualized volatility is almost 20%.

If time-varying parameters are drawn from its unconditional distribution (upper panel), results change. Predictive volatility starts low and increases up to 19% at the end of the 20-year horizon. Due to the combined effect of parameter uncertainty and time-variation in parameters, the term structure of risk is upward sloping for all horizons. At a 20-year horizon, predictive volatility is around 20% in both panels.

In both panels, the term structures deviate strongly from their time-constant counterpart. The latter is always strongly downward sloping. Apparently, if one ignores model instability, mean-reversion strongly dominates parameter uncertainty. Especially at the end of the sample, this gives a false sense of security. Stocks are much riskier if one takes time-variation in parameters into account.

Next, we calculate stock and bond weights for buy-and-hold investors who want to maximize expected power utility at time $t$ over terminal wealth at time $t + K$

$$\max_{w_s, w_b} E_t \left( \frac{W_{t+K}^{1-\gamma}}{1-\gamma} \right),$$

where relative risk aversion $\gamma = 5$, $w = (w_s, w_b)'$,

$$W_{t+K} = \prod_{j=1}^{K} (w' \exp(R_{tbill} + x_{t+j}) + (1 - w') \exp(R_{tbill}))$$

and where $t$ is an $(n - k) \times 1$ vector of ones. We consider investors who either (i) consider the time-varying or (ii) time-constant specification to calculate portfolio weights at both (a) the end of the sample and (b) when predictor variables are equal to their historical average. As above we draw time-varying parameters from its unconditional distribution in the latter case.

[Figure 4 about here.]

Figure 4 plots the results. Let us firstly consider the constant model. It is well-known - e.g. Campbell and Viceira (2002), Barberis (2000) - that the stock investment curve (plot of stock
weights versus investment horizon) is upward sloping in settings without model instability. We find similar results both at the end of the sample and at the historical average.

If we take time-variation in model parameters into account, the stock investment curve can however be strongly downward sloping. This is the case at the historical average. This is not a surprising result given the term structure of risk we showed in the previous figure. At the end of the sample, the stock investment curve is initially upward sloping, but becomes downward sloping for longer horizons. We conclude that this curve can have all kind of shapes if time-variation in parameters is taken into account. The bond investment curve is downward sloping in general.

If we compare the time-constant and time-varying specifications to each other, we see that portfolio weights differ a lot. For example, at the historical average, stock weights differ more than 40% at both short and long horizons. Ignoring model instability can therefore lead to huge investment mistakes.

In order to quantify investment mistakes, figure 5 plots the certainty equivalent returns (CERs) for buy-and-hold investors who either base their portfolio weights on the time-varying or time-constant specifications. In both cases, we evaluate the CERs using the time-varying specification. The certainty equivalent return for the time-varying specification is therefore by construction the highest. We consider the difference in CERs between both specifications. If ignoring time-variation leads to a considerable reduction in CER for an investor, we can conclude that ignoring model instability is economically costly.

[Figure 5 about here.]

The figure shows some interesting results. Firstly, the average CER across horizons is much higher at the historical average. Apparently, even at medium to long horizons, the unfavorable initial state leads to a considerable reduction in the attractiveness of financial markets. Secondly, at long horizons, ignoring time-variation can be really costly for investors. Differences can be as large as 5% per year. This is not surprising, since the previous figure showed that investment mistakes can be as large as 40%. Thirdly, if initial states deviate from their average values, it is also costly to ignore time-variation at short horizons. Even at the 1-month horizon, CERs are already reduced from more than 2% to 1%.

We conclude that it is economically important to take time variation into account. Ignoring time variation leads to a considerable underestimation of perceived risk, an overinvestment in the stock market and therefore a much lower performance. These effects are especially pronounced at long horizons and in case state variables deviate from their historical averages.

4.3 A closer look at the time-variation of model components

Our previous analysis shows the joint effect of time-variation in intercepts and slopes, time-variation in error volatility and error correlation and non-normality of the error terms on term structures of risk, portfolio weights and portfolio performance. In this section, we are interested in assessing the importance of the individual components of the time-varying model.
One way to do this would be to use statistical criteria such as posterior model probabilities. We do not pursue this alternative here because of two reasons. Firstly, it is well-known that posterior model probabilities are very sensitive to prior assumptions. Especially if one tries to choose prior distributions as uninformative as possible (as we do), the prior choice can have unintended consequences for model probabilities. Secondly, the portfolio weights and their expected utility are what matters ultimately for long-term investors.

Instead, we start with the most general model that nests all submodels and compare its portfolio weights to the portfolio weights of the submodels. These submodels are obtained by omitting some or all of the model components. We evaluate all portfolio weights using certainty equivalent returns calculated using the most general specification. The portfolio weights based on the most general model therefore have by definition the highest certainty equivalent return. What matters is whether other specifications lead to much lower certainty equivalent returns. If the omission of say time-varying slopes does not lead to a substantial loss of performance in a world where slopes indeed vary over time, then a risk-averse investor can safely ignore such an effect.

Table 4 in section 4.1 shows the posterior means and standard deviation of the most important hyperparameters of seven specifications that differ in whether $b$ is time-varying / time-constant, whether $\Omega$ is time-varying / time-constant and whether the error distributions contain fat tails / are Gaussian. Firstly, the table shows that ignoring stochastic volatility (column 1) leads to a large overestimation of the kurtosis of the error terms ($\tau^2$). This is in line with results in section 4.1 and table 2. Secondly, the persistence parameters for volatility are very robust across specifications. This suggests that the modeling of the time-variation in the error covariance matrix is not sensitive to the exact specification of intercepts and slopes. Thirdly, ignoring stochastic volatility implies a large underestimation of the persistence of the time-varying slopes and intercepts. For example, $A_{b,t(1,1)}$ is reduced from 0.91 to 0.32 if the error covariance matrix is constant. The reason is that ignoring stochastic volatility leads to many transitory movements in the time-varying slopes and intercepts due to outliers in periods when true volatility was actually very high. These transitory movements reduce the persistence of time-varying slopes and intercepts.

Table 5 shows the means and standard deviations of the posterior means of intercepts, slopes, error volatilities and error correlations over time. The table shows that the posterior means of time-varying intercepts and slopes are too volatile if stochastic volatility is ignored. This is caused by the presence of outliers in time-varying $b_t$ if the error covariance matrix is assumed constant. Furthermore, we clearly see that time-variation in intercepts and slopes is modest.

\[\text{Table 5 about here.}\]

\[\text{Figure 6 about here.}\]

\[\text{Table 5 shows the means and standard deviations of the posterior means of intercepts, slopes, error volatilities and error correlations over time. The table shows that the posterior means of time-varying intercepts and slopes are too volatile if stochastic volatility is ignored. This is caused by the presence of outliers in time-varying } b_t \text{ if the error covariance matrix is assumed constant. Furthermore, we clearly see that time-variation in intercepts and slopes is modest.}\]

\[^{10}\]The posterior means of the hyperparameters for the time constant model are not reported in the table, since they are all 0.
while time-variation in error volatility and error correlation is substantial across specifications. Finally, the table shows that unconditional means of the parameters vary quite a lot across specifications. For example, the incorporation of time-variation in $\Omega$ leads to large differences. If the latter is incorporated, periods with large volatility are underweighted when estimating unconditional means.

Figure 6 plots the annualized CERs for specifications that include some or all of the three model components (time-varying intercepts/slopes, time-varying covariance matrix, excess kurtosis). In all subpanels, the first specification is the time-constant model and the fourth specification is the time-varying model. The second and third specification respectively add the individual model component to the time-constant model or remove the individual model component from the fully time-varying model.

Let us firstly consider time-varying intercepts/slopes. The figure shows that it is very costly (at the end-of-the-sample) to add time-varying intercepts/slopes to the time-constant model. The incorporation of time-varying $b_t$’s, while ignoring time-varying volatility leads to a misspecified model. On the other hand, removing time-varying intercepts/slopes from the fully time-varying model is not costly. It only leads to a modest loss in CERs at all horizons.

Secondly, we consider the time-variation of the error covariance matrix. The figure shows that its inclusion always improves performance irrespective of the horizon. In fact, just including the time-varying error covariance matrix while ignoring time-varying $b_t$’s or excess kurtosis almost leads to the maximum performance. The omission of the time-variation in the error covariance matrix however turns out to be costly in almost all cases. The exception is the one-month horizon when predictor variables are equal to their historical average.

Finally, let us look at the incorporation of excess kurtosis. Clearly, adding excess kurtosis to the time-constant model only hurts performance. The time-constant model contains fat tails, because time-varying volatility is ignored. Apparently, allowing for non-normal distributions while ignoring stochastic volatility leads to a misspecified model and deteriorated performance. The removal of excess kurtosis from the fully time-varying model hardly has an affect on performance. This is not surprising, since previous sections show that the error distribution is close to normal once time-varying volatility is incorporated.

We conclude that the incorporation of time-varying intercepts/slopes and fat tails hardly has a positive effect on performance and can therefore be safely omitted. However, it is extremely important to incorporate a time-varying error covariance matrix. Its omission drastically reduces certainty equivalence returns.

4.4 The persistence parameter

The persistence of time-varying parameters plays a crucial role in the portfolio formation of long-term investors. This section analyzes the importance of the persistence by considering two alternative specifications. The first specification sets all persistence parameters equal to the same value. We estimate this specification to approximate regime-switching models, since all parameters have equal persistence in such models. Note that this pooled model is nested in our
most general model and is therefore a restricted version of this most general model. The aim of this section is to analyze the economic losses incurred by imposing these restrictions.

[Figure 7 about here.]

[Figure 8 about here.]

The second specification sets all persistence parameters equal to 1. We consider this random walk specification, because it is a popular way to model time-varying parameter models (e.g. Primiceri (2005)). Note that it is not nested in our most general specification. However, the results in the previous section show that there is no evidence at all that the persistence parameters for intercepts and slopes can be set to 1. The aim is again to assess the economic losses incurred if one uses this alternative specification.\(^{11}\)

[Figure 9 about here.]

[Figure 10 about here.]

Figure 7 reports the time-series of the posterior means of the intercepts and slope coefficients for three specifications: (i) the time-varying specification, (ii) the pooled specification and (iii) the random walk specification. The figure indicates that the three specifications lead to very different posterior means. Both the pooled as well as the random walk model are quite different from the time-varying model. Note that even for the random walk specification, the time-variation in the parameters is still relatively modest. The differences across specifications suggest that ”restricting” the persistence parameters to 1 or pooling the persistence parameters can lead to very different results.

Figure 8 plots time-series of the posterior means of error volatilities, correlations and parameters \(l_t\) for these three specifications. Remarkably, it is difficult to distinguish the three models with the naked eye. There is some difference in \(l_t,(3,1)\) and \(l_t,(3,2)\), but this does not lead to any noticeable differences in the posterior means of volatilities or correlations. We conclude that pooling persistence parameters or setting them to 1 is a viable alternative for the error covariance matrix.

Figure 9 plots the term structures of risk for the three specifications. Since the unconditional distribution of time-varying parameters is not defined for the random walk specification, we do not plot the term-structure at the historical average. The figure shows that restrictions lead to completely different term structures of risk. For horizons longer than 2 years, the term structures for the pooled and especially random walk specification increase extremely fast. At a 10 year horizon, the annualized predictive volatility is already more than 44%. For horizons longer than 10 years (not plotted), the predictive volatility reaches unrealistically high values. This suggests that restrictions should only be imposed for short horizons.

\(^{11}\)In unreported results, we also consider alternative specifications that nest the random walk specification. These specifications do no impose the existence of the long-run mean of the time-varying parameters. Results are in line with the results in this paper, i.e. no evidence for persistent variation in intercepts/slopes.
Figure 10 plots the portfolio weights for the three specifications. We again consider a buy-and-hold investor who maximizes expected power utility over final wealth with risk aversion parameter $\gamma = 5$. The figure shows that bond weights are very similar for all three models. However, the figure also shows that a risk-averse investor who either uses the pooled or random walk specification is much too conservative. Such an investor hardly invests in stocks at long horizons, since stock returns are much too risky in her eyes (see previous figure).

Finally, figure 11 plots the certainty equivalent returns for the three specifications. They are all evaluated under the time-varying model. Let us consider what happens at the historical average. The figure shows that an investor who uses a restricted specification hardly loses at very short horizons. Such an investor only starts to lose at horizons of 10 years or more. However, at the end of the sample, the situation is different. Here, investors lose at very short or very long horizons. The losses are still acceptable for medium horizons.

We conclude that it is economically important that persistence parameters are not restricted for different time-varying parameters. Restricting the parameters to be equal or imposing a random walk leads to an overestimation of risk, an underinvestment in the stock market and therefore to deteriorated performance at especially the longest horizons.

## 5 Robustness

Our main specification contains the dividend-to-price ratio as predictor for both stock and bond returns. One can argue that a different predictor for especially bond returns is more appropriate. In this section, we therefore consider the yield spread as an alternative predictor.

Figure 12 plots the posterior means of the intercepts and slopes over time. It reports results for the constant and time-varying model. The figure largely confirms the results of section 4.1. Firstly, there is modest time-variation in both intercepts and slopes in especially the stock return equation. The posterior mean of the yield spread coefficient varies between 0 and 0.40. Secondly, the time-variation is not very persistent and therefore not important for long-term investors. The diagonal elements of $A_{slope}$ are all around 0.85 which implies a half-life of less than half a year. Thirdly, the average coefficients differ quite a lot from the constant model. The difference is especially large for the lagged yield spread in the yield spread equation, 0.99 versus 0.96. This change in persistence is very important for long-term investors.

In order to understand the large change in persistence, we need to consider time-variation in the error covariance matrix. Figure 13 shows the posterior means of the error volatilities, error correlations and $l_t$ coefficients for both the time-constant and time-varying model. The figure confirms results that there is considerable time-variation in correlation and volatility. Firstly,
the most remarkable result is the large change in error volatility for the yield spread around 1980. Volatility increased to 0.015 which is much higher than its average value of 0.0028. This change is related to the change in the Fed policy when Volcker was appointed as chairman of the Federal Reserve in August 1979. The previous figure shows that ignoring time-variation in volatility leads to a large underestimation of the average persistence of the yield spread. Secondly, the correlations between the error in the yield spread and the errors in the stock and bond return equations vary a lot over time. The correlation for stocks is in general negative, but there are periods with positive correlation. This implies that there is mean-reversion in stocks on average, but that there are periods in which stocks show mean-aversion. The correlation for bonds is negative, but the strength of this correlation changes considerably over time. Long-term bonds strongly mean-revert in the 1930s and 1940s, but hardly mean-revert the 1960s and 1970s. Thirdly, the error volatility of stock and bond returns and the correlation between these errors is similar to figure 2. A change in predictor does not affect these parameters. This is not surprising given the low $R^2$ values of these regressions. Finally, the excess kurtosis in the error terms is considerably reduced if time-variation in volatility is incorporated. The posterior means of $\tau_1^2$, $\tau_2^2$ and $\tau_3^2$ are only 0.10, 0.00 and 0.44. Apparently, if one does not take the large change in volatility around 1980 into account, one would wrongly conclude that the yield spread contains extremely high excess kurtosis as in table 2. In other words, once the time-varying error covariance matrix is correctly modeled, there is hardly any excess kurtosis left.

We conclude that it is very important to take time-variation in the error covariance matrix into account. Ignoring the variation leads to a large underestimation of the persistence of the yield-spread and a large overestimation of the excess kurtosis of the error terms. The time-variation in intercepts and slopes is however not persistent enough for long-term investors. This confirms results of previous sections.

An alternative robustness check would have been to analyze whether the parameters governing the transition equations (10) are themselves time-varying. We could analyze this issue by extending the most general specification to allow for time-variation in these parameters and then test whether these parameters can be pooled over time. We do not pursue this alternative here because of several reasons. Firstly, we expect that such "time-varying time-variation" (if present at all) only has a second-order impact on portfolios. It seems unlikely that time-variation in the time-variation of intercepts and slopes is important given the results above. It could be important for the error covariance matrix, but it seems unlikely that it is as important as the direct time-variation in the error covariance matrix. Secondly, we would need to set-up transition equations for the intercepts, slopes and volatilities of the parameters in equations (10). This would increase the number of parameters considerably and would considerably increase the risk of overfitting. Finally, it would not solve the issue. One could wonder whether the parameters that govern the additional transition equations in the extended model are again time-varying etcetera.
6 Conclusion

This paper analyzes the effects of time-variation in model parameters on long-term investors. Our most general specification allows for time-varying intercepts and slope coefficients, time-varying error volatility and time-varying error correlation. It also allows for non-normal error distributions.

We find that the persistence of time-varying parameters plays a decisive role in the importance of different time-varying components. The time-variation in intercepts and slopes does not turn out to be persistent enough and is therefore not relevant for long-term investors and hardly has an impact on portfolio allocations. Time-variation in error correlations and especially error volatility is however very persistent and relevant for long-term investors due to the large impact on asset allocations. The normality assumption for the error term is valid as long as the error covariance matrix is allowed to vary over time. In case one ignores this time-variation, one would wrongly include that fat tails are important. Our preferred specification includes time-variation of the error covariance matrix, but ignores time-varying intercepts / slopes and ignores fat tails.

Our analysis can be extended in several directions. Firstly, we could consider a larger model that contains multiple predictor variables and multiple asset returns. Such a model contains a large number of parameters and its estimation therefore either requires restrictions or tight prior distributions. A second extension is to calculate fully dynamic strategies instead of straightforward buy-and-hold strategies. A full dynamic strategy includes learning about all parameters in the model. However, such a specification cannot be solved using the current state-of-the-art numerical techniques, because of the large number of state variables. An interesting alternative would be to ignore learning about hyperparameters, but to explicitly consider learning about the time-varying parameters. This limits the number of state variables considerable and might be feasible using techniques such as Brandt, Goyal, Santa-Clara, and Stroud (2005). Thirdly, since volatility affects the asset allocations of investors, it would be interesting to analyze whether it also affects expected returns directly by using volatility as a (latent) predictor of stock returns. This would however considerably complicate the MCMC algorithm.
A Posterior distribution and MCMC algorithm

In this appendix, we derive the posterior distributions and explain the MCMC algorithm to estimate the model.

Firstly, we consider how to draw time-varying parameters $b_t$, their unconditional means $\mu_b$, their transition matrix $A_b$ and the covariance matrix of its innovations $Q$. In order to do so, we reparameterize our model slightly

$$ y_t = X_t \mu_b + X_t \tilde{b}_t + u_t $$

$$ \ln \sigma^2_{t+1} = A_b \ln \sigma^2_t + \xi_t $$

$$ \mu_b = \mu_b $$

$$ \tilde{b}_{t+1} = A_b \tilde{b}_t + \eta_t. $$

The properties of the innovations are explained in section 3.1 and the initial conditions are specified in equations (14) and (22). The equations form a linear Gaussian state space model where the first two equations are the observation equations and the last two equations are the transition equations. We condition on (i) $A_b$, $Q$ (transition equation), (ii) $\lambda_t$, $\mu_t$, $\sigma_t$ $\forall t$ (the covariance matrix of $u_t$) and (iii) $A_s$, $S$, $\mu_s$ and $\rho$ (remaining terms second observation equation and correlation first and second observation equation). Therefore, we can use the standard Kalman filter - smoother technique to draw $\mu_b$ and $\tilde{b}_{t+1}$ $\forall t$ from $p(\mu_b, \tilde{b}_t | Y, A_b, A_s, Q, S, \lambda, l, \sigma, \rho)$. We use the Kalman filter - smoother technique explained in Durbin and Koopman (2002).

We include the second equation because $\xi_t$ is correlated with $u_t$ due to the leverage effect and cannot be ignored when drawing $\tilde{b}_t$ and $\mu_b$ conditional on (among others) $\sigma_t^2$. Therefore, we need to include it as an observation equation that is correlated to the first observation equation with correlation coefficient $\rho$. Note that in order to draw $\ln \sigma^2_t$ itself we also need to consider this same equation as a transition equation below.

If the initial condition (14) would not depend in a non-linear way on $A_b$ and $Q$, we could simply use an inverse Wishart distribution to draw $p(Q, \tilde{b}_t | Y, A_b, b)$ and a normal distribution to draw $p(A_b | Y, Q, b)$. Instead, we use a Metropolis-Hastings step where the inverse Wishart and normal distributions are used as proposal densities. For $Q$ the proposal density is

$$ \iota(Q^*) = iWishart \left( \sum_{t=1}^{T} (\eta_t \eta_t') + W_Q, T + df_Q \right) $$

and the acceptance probability is

$$ \alpha_b = \min \left\{ 1, \frac{|\Sigma^*_b|^{-\frac{1}{2}}}{|\Sigma_b|^{-\frac{1}{2}}} \exp \left( -\frac{1}{2} \tilde{b}_1' \Sigma^*_b \tilde{b}_1 + \frac{1}{2} \tilde{b}_1' \Sigma_b \tilde{b}_1 \right) \right\}, $$

where $\Sigma^*_b$ and $\Sigma_b$ are the unconditional covariance matrices - see the explanation below equation.
(14) - based on respectively the newly proposed draw $Q^*$ and the draw from the previous iteration.

Define $\tilde{B}$ as the $T \times nK$ matrix whose $t^{th}$ row is equal to $\tilde{b}_{t+1}$ and define $\tilde{B}_{-1}$ as the $T \times nK$ matrix whose $t^{th}$ row is equal to $\tilde{b}_t$. Let $a_b$ be the diagonal of $A_b$. Its proposal density is

$$\iota(a_b^*) = N(\tilde{a}_b, M_b^{* -1} ),$$

where

$$M_b^* = X_b' (Q^{-1} \otimes I_T) X_b + v_b^{-1} I_{nK},$$

$$\tilde{a}_b = M_b^{-1} \left( X_b' (Q^{-1} \otimes I_T) y_b + \iota_{nK} v_b^{-1} m_b \right),$$

where $y_b$ vertically stacks the columns of $B$, where

$$X_b = \begin{pmatrix} \tilde{b}_{1,-1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \tilde{b}_{nK,-1} \end{pmatrix}$$

and where $\tilde{b}_{i,-1}$ is the $i^{th}$ column of $B_{-1}$. This is similar to a GLS regression. The acceptance probability is again $\alpha_b$ with $\Sigma_b^*$ depending on the newly proposed draw $A_b^*$. 

Secondly, we look at drawing time-varying parameters $l_t$, their unconditional means $\mu_l$, transition matrix $A_l$ and the covariance matrix of the innovations $R$. The system of equations can be rewritten as

$$L_t(y_t - X_l b_t) \equiv L_t \hat{y}_t = \Lambda_t \Sigma_t e_t.$$

We condition on $b_t$, $\Sigma_t$ and $\Lambda_t \forall t$ and can therefore treat them as given in this step. Since matrix $L_t$ is a lower triangular matrix with ones on the diagonal, we can rewrite the previous equation as

$$\hat{y}_t = T_t l_t + \Lambda_t \Sigma_t e_t, $$

where

$$T_t = \begin{pmatrix} 0 & \cdots & \cdots & 0 \\ -\hat{y}_{1,t} & 0 & \cdots & 0 \\ 0 & -\hat{y}_{[1:2],t} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & -\hat{y}_{[1:n-1],t} \end{pmatrix}.$$ 

Since the dependent variable of the observation equation - $\hat{y}_t$ - also occurs on the right-hand-side (RHS) in $T_t$, this system has a nonlinear Gaussian state space representation. However, since (i) we impose that $R$ is blockdiagonal (ii) the system has a triangular structure and (iii) the dependent variable in one equation does not show up on the RHS of the same equation, we
can apply the Kalman filter - smoother technique equation by equation. Hence, for equation $j$ with $j = 2, ..., n$, we consider the following linear Gaussian state space model

$$
\begin{align*}
\hat{y}_{t,j} &= -\hat{y}_{t,[1:j-1]|j} - \hat{y}_{t,[1:j-1]|t} + \tilde{\lambda}_{t,j}^{2} \sigma_{j,t} e_{t,j} \\
\ln \sigma_{j,t+1}^{2} &= A_{l,(j,j)} \ln \sigma_{j,t} + \xi_{t,j} \\
\mu_{t,(j)} &= \mu_{t,(j)} \\
\tilde{t}_{t,j+1} &= A_{l,(j,j)} \tilde{t}_{t,j} + \zeta_{t},
\end{align*}
$$

(47)

where $\{j\}$ refers to the elements of the vectors/matrices that belong to the $j^{th}$ equation. The properties of the error terms are explained in section 3.1 and the initial conditions are given in equations (15) and (23). The first two equations are the observations equations and the last two are the transition equations. As above, the second equation is included, because the correlation between $e_{t,j}$ and $\xi_{t,j}$ is equal to $\rho_{j}$. We can use the Kalman filter - smoother technique to draw $\mu_{t,(j)}$ and $\tilde{t}_{t,(j)}$ $\forall \ t$ from $p(\mu_{t,(j)}, \tilde{t}_{t,(j)}|Y, A_{s,(j,j)}, S_{j,j}, \sigma_{j}, \lambda_{j}, A_{l,(j,j)}, R_{(j,j)})$.

We cannot simply use Gibbs steps to draw $R_{(j)}$ and $A_{l,(j)}$ for $j = 2, ..., n$, because the unconditional variance $\Sigma_{l,(j)}$ depends in a non-linear way on these parameters. We use Metropolis-Hastings steps instead where we use inverse Wishart and normal distributions as proposal densities for $p(R_{2:3,2:3}|Y, A_{l,(3)}, \tilde{t}_{(3)})$, $p(A_{l,(2)}|Y, R_{1:1}, \tilde{t}_{(2)})$ and $p(A_{l,(3)}|Y, R_{2:3,2:3}, \tilde{t}_{(3)})$. Since these steps are almost similar to drawing $A_{h}$ and $Q$ above, we do not explicitly write them down. We do not need a Metropolis-Hastings step when drawing $R_{1,1}$ for $j = 2$. In this case, we can use a Gibbs step by drawing from the inverse gamma distribution

$$
p(R_{1,1}|Y, A_{l,(1,1)}, \tilde{t}_{(2)}) = \text{iGamma} \left( \left[ (1 - A_{l,(1,1)})^{2} \right]^{T} + \sum_{t=1}^{T} (\zeta_{t}^{2})^{|} + W_{R1}, T + 1 + df_{R1} \right).
$$

(48)

Thirdly, we explain how to simulate the time-varying parameters $\ln \sigma_{j}^{2}$ and $\lambda_{t} \ \forall t$, the unconditional mean $\mu_{s}$, transition matrix $A_{s}$, covariance matrix $S$ and correlation coefficient $\rho$. We rewrite our model slightly

$$
L_{t}(y_{t} - X_{t}b_{t}) \equiv y_{t}^{*} = \Lambda_{t} \Sigma_{t} e_{t}.
$$

(49)

We condition on $l_{t}$ and $b_{t}$ and hence treat them as given. Note that we can treat the different equations separately, since (i) the elements of $e_{t}$ are independent of each other and (ii) we impose a diagonal structure for covariance matrix $S$.

The observation equations are non-linear in the diagonal elements of $\Lambda_{t}$ and $\Sigma_{t}$ which means that we cannot use the linear Gaussian state space model without any further modification. Following Omori, Chib, Shephard, and Nakajima (2007) we consider the following transformation of the dependent variable instead

$$
y_{t,j}^{**} = \ln y_{t,j}^{*2}, \forall t, j.
$$

(50)
For $j = 1, \ldots, n$ this gives the following state space model

$$y^{*\ast}_{t,j} = \mu_{s,j} + \ln\sigma^2_{t,j} + \ln(\lambda_{t,j}) + \ln(e^2_{t,j})$$

$$\mu_{s,j} = \mu_{s,j}$$

$$\ln\sigma^2_{t+1,j} = A_{s,(j,j)}\ln\sigma^2_{t,j} + \xi_{t,j}$$

$$\ln(\lambda_{t+1,j}) = -(1/2)\tau_j^2 + u_{r,j},$$

where

$$u_{r,j} \sim N(0, \tau_j^2)$$

is independent of the other error terms. The properties of the other innovations are given in section 3.1 and the initial conditions are specified in equations (16) and (24). The first equation is the observation equation and the remaining equations are the transition equations. The error in the observation equation - $\ln(e^2_{t,j})$ - and the innovation in the transition equation - $\xi_{t,j}$ - are dependent, since the correlation between $e_{t,j}$ and $\xi_{t,j}$ is $\rho_j$.

The state space system is linear, but non-Gaussian, since the error in the observation equation has a log $\chi^2$ distribution. In order to be able to use linear Gaussian state space techniques, Omori, Chib, Shephard, and Nakajima (2007) propose to approximate the log chi-squared distribution using a mixture of 10 normal distributions. Their method is an extension of Kim, Shephard, and Chib (1998) by allowing for the leverage effect, i.e. dependence between $\ln(e^2_{t,j})$ and $\xi_{t,j}$. It allows us to draw from $p(\ln\sigma^2_j, \ln(\lambda_j)|Y, b, l, A_{s,(j,j)}, S_{j,j}, \tau_j^2, \rho_j)$. Please refer to Omori, Chib, Shephard, and Nakajima (2007) for more details.

In order to draw $\theta_j$ and $\psi_j$ (which we transform to $\rho_j$ and $S_{j,j}$), we use results in Jacquier, Polson, and Rossi (2004). Let $r_{t,j} = (e_{t,j}, \xi_{t,j})$ for equation $j$. The posterior distribution of $\Sigma_j^*$ is proportional to

$$p(\Sigma^*_j|Y, A_{s,(j,j)}, \ln\sigma^2_j) \propto p(\Sigma^*_j)p(\ln\sigma^2_j|A_{s,(j,j)}, S_{j,j})|\Sigma^*_j|^{-T/2} \exp\left(-\frac{1}{2} \text{tr} \left(\Sigma_j^* - 1UU_j\right)\right),$$

where $UU_j = \sum_r r_{t,j}r_{t,j}'$.

Define $a_{(k,l),j}$ as the $(k, l)^{th}$ element of $UU_j$. Furthermore, let $a_{22,1,j} = a_{(2,2),j} - a_{(1,2),j}^2/a_{(1,1),j}$ and let $\tilde{\psi}^j = a_{(1,2),j}/a_{(1,1),j}$. Ignoring the initial condition, it is easy to show (see Jacquier, Polson, and Rossi (2004)) that

$$p(\psi_j|Y, \theta_j, b, \lambda_{t,j}, \ln\sigma^2_j, A_{s,(j,j)}) = N(\tilde{\psi}_j, \frac{\theta_j}{a_{(1,1),j} + p})$$

$$p(\theta_j|Y, b, \lambda_{t,j}, \ln\sigma^2_j, A_{s,(j,j)}) = i\Gamma(a_{22,1,j} + W_{(a, j), T + df_\theta}),$$

where

$$\tilde{\psi}_j = \frac{a_{(1,1),j}\psi^j}{a_{(1,1),j} + p}$$

These are the conditional distributions used in Jacquier, Polson, and Rossi (2004).
are not equal to the conditional posteriors in our setting due to the presence of the initial conditions. We use these two equations as proposal densities in a Metropolis-Hastings step with the following acceptance probability

$$
\alpha_s = \min \left\{ 1, \frac{|\Sigma_s^*|^{-\frac{3}{2}}}{|\Sigma_s|^{-\frac{3}{2}}} \exp \left( -\frac{1}{2} \ln \sigma_{i,j}^2 \Sigma_s^* \ln \sigma_{i,j}^2 + \frac{1}{2} \ln \sigma_{i,j}^2 \Sigma_s \ln \sigma_{i,j}^2 \right) \right\},
$$

(57)

where $\Sigma_{s,(j,j)}^*$ depends on the newly drawn $\psi_j^*$ and $\theta_j^*$.

The step to draw $A_{s,(j,j)}$ for equations $j = 1, \ldots, n$ is a relatively straightforward Metropolis-Hastings step. We need to take the correlation between $\xi_{t,j}$ and $e_{t,j}$ into account. Therefore, we obtain the following auxiliary equation for $\tilde{\ln} \sigma_{t+1,j}^2$

$$
\tilde{\ln} \sigma_{t+1,j}^2 \equiv \tilde{\ln} \sigma_{t,j}^2 - \Delta \times e_t = A_{s,(j,j)} \tilde{\ln} \sigma_{t,j}^2 + \xi_t^*,
$$

(58)

where innovation $\xi_t^* \sim N(0, S_{j,j}(1 - \rho^2_j))$ (due to the conditioning) and where

$$
\Delta = \rho_j \sqrt{S_{j,j}}.
$$

(59)

It is straightforward to show that we can use the following proposal density for $A_{s,(j,j)}^*$

$$
t(A_{s,(j,j)}^*) = N(\hat{a}_j, M_j^{-1})
$$

(60)

with

$$
M_j = (S_{j,j}(1 - \rho^2_j))^{-1} \left( \sum_{t=1}^{T} (\tilde{\ln} \sigma_{t,j}^2)^2 \right) + v_s^{-1}
$$

(61)

and

$$
\hat{a}_j = M_j^{-1} \left( (S_{j,j}(1 - \rho^2_j))^{-1} \sum_{t=1}^{T} (\tilde{\ln} \sigma_{t,j}^2 \tilde{\ln} \sigma_{t+1,j}^2) + v_s^{-1} m_s \right)
$$

(62)

and the same acceptance probability as in equation (57) where $\Sigma_{s,(j,j)}^*$ depends on the newly drawn value for $A_{s,(j,j)}^*$.

Finally, we look at drawing $\tau_j^2$. The kernel of the conditional posterior distribution for $\tau_j^2$ is

$$
p(\tau_j^2 | Y, \lambda_j) \propto (\tau_j^2)^{-\frac{T+1.5}{2}} \exp \left( -\frac{\tau_j^2}{2} \right) \prod_{t=1}^{T} \exp \left( \frac{1}{2\tau_j^2} (\ln \lambda_{t,j} + 0.5\tau_j^2)^2 \right).
$$

(63)

This is a non-standard distribution, but we simulate from this distribution for $j = 1, \ldots, n$ using a Metropolis-Hastings step. More precisely, we simulate $\ln \tau_j^2$ by finding the mode $m_j$ of the above posterior distribution in every iteration and by subsequently using a student-t distribution as proposal density with mean equal to the log of the mode, variance equal to 1.1 times the negative inverse of the hessian matrix of the log kernel at the mode and degrees of freedom equal to 8. The acceptance probability is calculated in the usual way using both the proposal and the
kernel (taking the Jacobian of the transformation into account), refer to for example Bauwens, Lubrano, and Richard (1999), page 89. Finally, we transform $\ln \tau^2_j$ to obtain $\tau^2_j$.\footnote{The acceptance probability using $\ln \tau^2_j$ is approximately 97%. If we would simulate $\tau^2_j$ directly, the acceptance probability would decrease to 67%.}

The acceptance probabilities for all Metropolis-Hastings steps are larger than 80% in all cases. In the empirical section, we retain 10,000 iterations after a burn-in period of 5,000 iterations. Increasing the number of iterations does not significantly impact results. We draw 10 path asset return paths per iteration, 100,000 paths in total, to calculate predictive distributions and portfolio weights.
References


**List of Figures**

1. Time-series of the posterior means of intercepts and slopes ........................................ 32
2. Time-series of the posterior means of error volatilities, correlations and $l_t$ .......... 33
3. Annualized predictive standard deviation of excess log stock returns: time-varying and time-constant specifications ................................................................. 34
4. Optimal stock and bond weights: time-varying and time-constant specifications .... 35
5. Annualized certainty equivalent returns: time-varying and time-constant specifi-
cations .................................................................................................................................. 36
6. Annualized certainty equivalent returns: individual model components .......... 37
7. Time-series of the posterior means of the intercepts and slopes for the time-
varying, pooled and random walk specifications ................................................................. 38
8. Time-series of the posterior means of error volatilities, correlations and $l_t$ for the time-varying, pooled and random walk specifications ................................. 39
9. Annualized predictive standard deviation of excess log stock returns: time-varying, pooled and random walk specifications ................................................................. 40
10. Optimal stock and bond weights: time-varying, pooled and random walk specifi-
cations .................................................................................................................................. 41
11. Annualized certainty equivalent returns: time-varying, pooled and random walk specifi-
cations .................................................................................................................................. 42
12. Time-series of the posterior means of the intercepts and slopes for $Y_{spr}$ model .... 43
13. Time-series of the posterior means of error volatilities, correlations and $l_t$ for the $Y_{spr}$ model ................................................................. 44
This figure shows the (smoothed) posterior means of intercepts and slopes for the time-varying and the time-constant model. Results are based on 10,000 retained draws. Note that the equations are given in the different rows.
Figure 2: Time-series of the posterior means of error volatilities, correlations and $l_t$

This figure shows the (smoothed) posterior means of the standard deviation of the residuals (diagonal), the correlation between the residuals (above the diagonal) and the $l_t$ coefficients (below the diagonal) for the time-varying and time-constant specifications. Results are based on 10,000 retained draws. Note that the equations are given in the different rows.
The figure shows the annualized predictive standard deviation of excess log stock returns for the time-varying and time-constant specifications. Time-varying parameters are either drawn from its unconditional distribution (upper panel) or from its posterior distribution at the end of the sample (lower panel). The predictor variable is either set to its historical average (upper panel) or to its end-of-sample value (lower panel). Results are based on 100,000 retained draws.
The figure shows the optimal stock weights (first column) and optimal bond weights (second column) for a buy-and-hold investor with $\gamma = 5$ who either uses the time-varying specification or the time-constant specification. Time-varying parameters are either drawn from its unconditional distribution (first row) or from its posterior distribution at the end of the sample (second row). The predictor variable is either equal to the historical average (first row) or to the end-of-sample value (second row). Results are based on 100,000 retained draws.
This figure shows the annualized certainty equivalent returns for buy-and-hold investors with $\gamma = 5$ who either base their portfolio weights on the time-varying specification or the time-constant specifications. The certainty equivalent returns are calculated using the time-varying specification. Time-varying parameters are either drawn from its unconditional distribution (first row) or from its posterior distribution at the end of the sample (second row) and the predictor variable is either equal to its historical average (first row) or to its end-of-sample value (second row). Results are based on 100,000 retained draws.
This figure shows the annualized certainty equivalent return for a buy-and-hold investor with $\gamma = 5$ who uses several specifications to calculate portfolio weights. These specifications differ in whether they allow for time-varying intercepts/slopes, a time-varying covariance matrix and excess kurtosis. The certainty equivalent returns are calculated using the time-varying specification. We report results for the 1 month horizon, the 10 years horizon and the 20 years horizon. Time-varying parameters are either drawn from its posterior distribution at the end of the sample (first row) or its unconditional distribution (second row). Results are based on 100,000 retained draws. The first and fourth column in every subpanel report the time-constant model and the time-varying model. The second column shows a specification where either the time-varying $b_t$, time-varying covariance matrix or excess kurtosis is added to the time-constant model. The third column reports the results for a setting in which either time-varying $b_t$’s, time-varying covariance matrix or excess kurtosis is removed from the time-varying model.
The figure shows the (smoothed) posterior means of the intercept and slopes for the time-varying model, the model in which the persistence parameters are equal for all time-varying parameters and the random walk specification. Results are based on 10,000 retained draws. Note that the equations are given in different rows.
The figure shows the (smoothed) posterior means of the standard deviation of the residuals (diagonal), the correlation between the residuals (above the diagonal) and the $l_t$ coefficients (below the diagonal) for the time-varying model, the model in which the persistence parameters are equal for all time-varying parameters and finally the random walk specification. Results are based on 10,000 retained draws. Note that equations are given in the different rows.
The figure shows the annualized predictive standard deviation of excess log stock returns for the time-varying specification, the specification in which the persistence parameters are pooled for all time-varying parameters and finally the random walk specification. Time-varying parameters are either drawn from its unconditional distribution (upper panel) or from its posterior distribution at the end of the sample (lower panel). The predictor variable is either equal to its historical average (first row) or to its end-of-sample value (second row). Results are based on 100,000 retained draws.
The figure shows the optimal stock weights (first column) and optimal bond weights (second column) for a buy-and-hold investor with $\gamma = 5$ who bases her portfolio weights on either the time-varying specification, the specification in which the persistence parameters are pooled for all time-varying parameters and finally the random walk specification. Time-varying parameters are either drawn from its unconditional distribution (first row) or from its posterior distribution at the end of the sample (second row). The predictor variable is either equal to its historical average (first row) or to its end-of-sample value (second row). Results are based on 100,000 retained draws.
The figure shows the annualized certainty equivalent return for a buy-and-hold investor with $\gamma = 5$ who either bases her portfolio weights on the time-varying specification, the specification in which the persistence parameters are pooled for all time-varying parameters and finally the random walk specification. All certainty equivalent returns are calculated using the time-varying specification. Time-varying parameters are either drawn from its unconditional distribution (first row) or from its posterior distribution at the end of the sample (second row). The predictor variable is either equal to its historical average (first row) or to its end-of-sample value (second row). Results are based on 100,000 retained draws.
Figure 12: **Time-series of the posterior means of the intercepts and slopes for \( Y_{spr} \) model**

The figure shows the (smoothed) posterior means of the intercept and slopes for the time-varying and time-constant \( Y_{spr} \) - model. Results are based on 10,000 retained draws. Note that the equations are given in the different rows.
Figure 13: Time-series of the posterior means of error volatilities, correlations and $l_t$ for the $Y_{spr}$ model

The figure shows the (smoothed) posterior means of the standard deviation of the residuals (diagonal), the correlation between the residuals (above the diagonal) and the $l_t$ coefficients (below the diagonal) for the time-varying and time-constant $Y_{spr}$ - specifications. Results are based on 10,000 retained draws. Note that the equations are given in the different rows.
List of Tables

1. Summary Statistics of the monthly data-set .............................................. 46
2. Preliminary analysis ................................................................................ 47
3. Moments of the normal-lognormal distribution ........................................ 48
4. Posterior means and standard deviations of hyperparameters ................. 49
5. Means and standard deviations of the posterior means of the time-varying parameters .......................................................... 50
Table 1: **Summary Statistics of the monthly data-set**

This table reports the means, standard deviations, minima, maxima, AR(1) coefficients, Skewness, Kurtosis and Sharpe ratios for excess stock returns ($X_s$), excess bond returns ($X_b$), the dividend-to-price ratio (DP) and the yield spread ($Y_{spr}$). The data set is monthly and starts in December 1926 and ends in December 2008. Percentages are given as fractions.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std</th>
<th>Min</th>
<th>Max</th>
<th>AR(1)</th>
<th>Skewness</th>
<th>Kurtosis</th>
<th>Sharpe</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_s$</td>
<td>0.0014</td>
<td>0.0557</td>
<td>-0.3391</td>
<td>0.3471</td>
<td>0.0904</td>
<td>-0.4007</td>
<td>10.8715</td>
<td>0.0793</td>
</tr>
<tr>
<td>$X_b$</td>
<td>0.0015</td>
<td>0.0231</td>
<td>-0.1041</td>
<td>0.1342</td>
<td>0.0556</td>
<td>0.3253</td>
<td>7.4090</td>
<td>0.0640</td>
</tr>
<tr>
<td>DP</td>
<td>-3.3399</td>
<td>0.4622</td>
<td>-4.5074</td>
<td>-1.6851</td>
<td>0.9924</td>
<td>-0.2759</td>
<td>3.2983</td>
<td></td>
</tr>
<tr>
<td>$Y_{spr}$</td>
<td>0.0153</td>
<td>0.0122</td>
<td>-0.0319</td>
<td>0.0438</td>
<td>0.9600</td>
<td>-0.2259</td>
<td>3.0648</td>
<td></td>
</tr>
</tbody>
</table>
Table 2: Preliminary analysis

This table shows the OLS estimates, standard errors and covariance matrix for respectively the model with DP (panel A) and the model with $Y_{spr}$ (panel B) as predictor. The first subpanel gives the parameter estimates, the standard errors in brackets and the skewness and kurtosis of the error terms. The second subpanel shows the estimates for the covariance matrix of the residual. The diagonal indicates the standard errors of the residuals whereas the off-diagonal elements are the correlations. Note that the different equations are given in different rows and that $c$ indicates the constant.

<table>
<thead>
<tr>
<th>Panel A1: Est. DP model</th>
<th></th>
<th></th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_s$</td>
<td>0.0270</td>
<td>0.0068</td>
<td>-0.5262</td>
<td>10.7775</td>
</tr>
<tr>
<td>$X_b$</td>
<td>0.0031</td>
<td>0.0005</td>
<td>0.3293</td>
<td>7.3907</td>
</tr>
<tr>
<td>DP</td>
<td>-0.0258</td>
<td>0.9924</td>
<td>0.2568</td>
<td>11.4675</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel A2: Cov. DP model</th>
<th>$X_s$</th>
<th>$X_b$</th>
<th>DP</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_s$</td>
<td>0.0556</td>
<td>0.1227</td>
<td>-0.9860</td>
</tr>
<tr>
<td>$X_b$</td>
<td>0.1227</td>
<td>0.0563</td>
<td>-0.1253</td>
</tr>
<tr>
<td>DP</td>
<td>-0.9860</td>
<td>-0.1253</td>
<td>0.0563</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B1: Est. $Y_{spr}$ model</th>
<th></th>
<th>$Y_{spr}$</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_s$</td>
<td>0.0023</td>
<td>0.1373</td>
<td>-0.4238</td>
<td>10.8776</td>
</tr>
<tr>
<td>$X_b$</td>
<td>-0.0026</td>
<td>0.2640</td>
<td>0.3516</td>
<td>7.6976</td>
</tr>
<tr>
<td>$Y_{spr}$</td>
<td>0.0006</td>
<td>0.9600</td>
<td>0.5155</td>
<td>25.1347</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B2: Cov. $Y_{spr}$ model</th>
<th>$X_s$</th>
<th>$X_b$</th>
<th>$Y_{spr}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_s$</td>
<td>0.0557</td>
<td>0.1201</td>
<td>-0.0185</td>
</tr>
<tr>
<td>$X_b$</td>
<td>0.1201</td>
<td>0.0228</td>
<td>-0.2774</td>
</tr>
<tr>
<td>$Y_{spr}$</td>
<td>-0.0185</td>
<td>-0.2774</td>
<td>0.0034</td>
</tr>
</tbody>
</table>
Table 3: Moments of the normal-lognormal distribution

This table reports the first four moments of the normal-lognormal distribution $\lambda_t e_t$ with $\ln \lambda_t \sim N(-1/2\tau^2, \tau^2)$ and $e_t \sim N(0,1)$ for different values of $\tau^2$.

<table>
<thead>
<tr>
<th>$\tau^2$</th>
<th>0.0</th>
<th>0.2</th>
<th>0.4</th>
<th>0.6</th>
<th>0.8</th>
<th>1.0</th>
<th>1.2</th>
<th>1.4</th>
<th>1.6</th>
<th>1.8</th>
<th>2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Std</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Skewness</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>3.0</td>
<td>3.7</td>
<td>4.5</td>
<td>5.5</td>
<td>6.6</td>
<td>8.1</td>
<td>9.2</td>
<td>12.0</td>
<td>14.5</td>
<td>18.2</td>
<td>22.8</td>
</tr>
</tbody>
</table>
Table 4: Posterior means and standard deviations of hyperparameters

This table shows the posterior means and standard deviations (between brackets) of the most important hyperparameters of seven specifications. The specifications differ in whether $b$ is time-varying (1) / time-constant (0), whether $\Omega$ is time-varying (1) / time-constant (0) and finally whether error terms have fat tails (1) / are gaussian (0). Matrices $A_b$, $A_s$ and $A_l$ are the transition matrices for $b_t$, $\ln \sigma^2_t$ and $l_t$. Vectors $\rho$ and $\tau^2$ are respectively the correlation between error term $e_t$ and $\xi_t$, and a kurtosis measure for $e_t$.

<table>
<thead>
<tr>
<th></th>
<th>TV $b$</th>
<th>TV $\Omega$</th>
<th>Kurt</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$A_{b,(1,1)}$</td>
<td>0.3110</td>
<td>0.3269</td>
<td>0.9153</td>
<td>0.9109</td>
<td>0.0480</td>
<td>0.0483</td>
<td>0.0318</td>
<td>0.0347</td>
<td></td>
</tr>
<tr>
<td>$A_{b,(2,2)}$</td>
<td>0.6703</td>
<td>0.6623</td>
<td>0.8804</td>
<td>0.8837</td>
<td>0.0884</td>
<td>0.1051</td>
<td>0.0319</td>
<td>0.0321</td>
<td></td>
</tr>
<tr>
<td>$A_{b,(3,3)}$</td>
<td>0.8112</td>
<td>0.9353</td>
<td>0.8605</td>
<td>0.8637</td>
<td>0.1354</td>
<td>0.1107</td>
<td>0.1416</td>
<td>0.1213</td>
<td></td>
</tr>
<tr>
<td>$A_{b,(4,4)}$</td>
<td>0.7764</td>
<td>0.8337</td>
<td>0.8383</td>
<td>0.8779</td>
<td>0.1627</td>
<td>0.1519</td>
<td>0.1463</td>
<td>0.1310</td>
<td></td>
</tr>
<tr>
<td>$A_{b,(5,5)}$</td>
<td>0.9112</td>
<td>0.8853</td>
<td>0.9411</td>
<td>0.9436</td>
<td>0.0178</td>
<td>0.0205</td>
<td>0.0163</td>
<td>0.0144</td>
<td></td>
</tr>
<tr>
<td>$A_{b,(6,6)}$</td>
<td>0.4837</td>
<td>0.5674</td>
<td>0.8973</td>
<td>0.9021</td>
<td>0.0850</td>
<td>0.0784</td>
<td>0.0555</td>
<td>0.0540</td>
<td></td>
</tr>
<tr>
<td>$A_{s,(1,1)}$</td>
<td>0.9538</td>
<td>0.9589</td>
<td>0.9495</td>
<td>0.9555</td>
<td>0.0184</td>
<td>0.0175</td>
<td>0.0183</td>
<td>0.0169</td>
<td></td>
</tr>
<tr>
<td>$A_{s,(2,2)}$</td>
<td>0.9784</td>
<td>0.9788</td>
<td>0.9773</td>
<td>0.9808</td>
<td>0.0090</td>
<td>0.0090</td>
<td>0.0091</td>
<td>0.0085</td>
<td></td>
</tr>
<tr>
<td>$A_{s,(3,3)}$</td>
<td>0.9865</td>
<td>0.9860</td>
<td>0.9984</td>
<td>0.9985</td>
<td>0.0067</td>
<td>0.0065</td>
<td>0.0016</td>
<td>0.0015</td>
<td></td>
</tr>
<tr>
<td>$A_{l,(1,1)}$</td>
<td>0.9964</td>
<td>0.9959</td>
<td>0.9962</td>
<td>0.9960</td>
<td>0.0028</td>
<td>0.0033</td>
<td>0.0030</td>
<td>0.0032</td>
<td></td>
</tr>
<tr>
<td>$A_{l,(2,2)}$</td>
<td>0.8168</td>
<td>0.8069</td>
<td>0.8352</td>
<td>0.8506</td>
<td>0.1328</td>
<td>0.1313</td>
<td>0.1183</td>
<td>0.1292</td>
<td></td>
</tr>
<tr>
<td>$A_{l,(3,3)}$</td>
<td>0.7925</td>
<td>0.7733</td>
<td>0.8182</td>
<td>0.7599</td>
<td>0.1276</td>
<td>0.1567</td>
<td>0.1361</td>
<td>0.1325</td>
<td></td>
</tr>
<tr>
<td>$\tau^2_1$</td>
<td>0.8802</td>
<td>0.0592</td>
<td>0.9145</td>
<td>0.1269</td>
<td>0.1368</td>
<td>0.0506</td>
<td>0.1422</td>
<td>0.0702</td>
<td></td>
</tr>
<tr>
<td>$\tau^2_2$</td>
<td>1.3335</td>
<td>0.0445</td>
<td>1.6542</td>
<td>0.1626</td>
<td>0.2011</td>
<td>0.0723</td>
<td>0.2569</td>
<td>0.1147</td>
<td></td>
</tr>
<tr>
<td>$\tau^2_3$</td>
<td>1.8091</td>
<td>0.0003</td>
<td>0.2506</td>
<td>0.0750</td>
<td>0.2191</td>
<td>0.0004</td>
<td>0.3696</td>
<td>0.0018</td>
<td></td>
</tr>
<tr>
<td>$\rho_1$</td>
<td>-0.3249</td>
<td>-0.3513</td>
<td>-0.3629</td>
<td>-0.4266</td>
<td>-0.3249</td>
<td>-0.3513</td>
<td>-0.3629</td>
<td>-0.4266</td>
<td></td>
</tr>
<tr>
<td>$\rho_2$</td>
<td>-0.0856</td>
<td>-0.1024</td>
<td>-0.1108</td>
<td>-0.1532</td>
<td>-0.0902</td>
<td>-0.0981</td>
<td>-0.1094</td>
<td>-0.1250</td>
<td></td>
</tr>
<tr>
<td>$\rho_3$</td>
<td>0.0714</td>
<td>0.0711</td>
<td>-0.5311</td>
<td>-0.4787</td>
<td>0.0511</td>
<td>0.0530</td>
<td>0.2148</td>
<td>0.2351</td>
<td></td>
</tr>
</tbody>
</table>
Table 5: Means and standard deviations of the posterior means of the time-varying parameters

This table plots the means and standard deviations of the posterior means of the time-varying parameters over time for eight specifications. The specifications differ in whether \( b \) is time-varying (1) / time-constant (0), whether \( \Omega \) is time-varying (1) / time-constant (0) and finally whether error terms have fat tails (1) / are gaussian (0). The first part of the table shows the means and standard deviations of the posterior mean of \( b_t \) over time. The second part gives the means and standard deviations of the posterior mean of \( \Omega_t \) over time. The diagonal elements show the moments of the error volatilities and the off-diagonal elements show the moments of the error correlations.

| TV \( b \) | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 |
| TV \( \Omega \) | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 |
| Kurt | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 |
| \( b_{1,1} \) | 0.0368 | 0.0324 | 0.0280 | 0.0273 | 0.0405 | 0.0337 | 0.0454 | 0.0453 |
| | (0.0085) | (0.0070) | (0.0042) | (0.0038) | | | | |
| \( b_{1,2} \) | 0.0094 | 0.0072 | 0.0063 | 0.0061 | 0.0105 | 0.0076 | 0.0115 | 0.0114 |
| | (0.0009) | (0.0017) | (0.0011) | (0.0011) | | | | |
| \( b_{2,1} \) | 0.0046 | 0.0027 | 0.0070 | 0.0070 | 0.0092 | 0.0144 | 0.0078 | 0.0075 |
| | (0.0006) | (0.0032) | (0.0008) | (0.0008) | | | | |
| \( b_{2,2} \) | 0.0009 | 0.0003 | 0.0019 | 0.0019 | 0.0023 | 0.0038 | 0.0020 | 0.0018 |
| | (0.0004) | (0.0004) | (0.0002) | (0.0003) | | | | |
| \( b_{3,1} \) | -0.0349 | -0.0254 | -0.0217 | -0.0210 | -0.0396 | -0.0384 | -0.0562 | -0.0577 |
| | (0.0114) | (0.0137) | (0.0106) | (0.0108) | | | | |
| \( b_{3,2} \) | 0.9900 | 0.9939 | 0.9949 | 0.9951 | 0.9886 | 0.9898 | 0.9841 | 0.9838 |
| | (0.0145) | (0.0141) | (0.0125) | (0.0125) | | | | |
| \( \Omega_{1,1} \) | 0.0505 | 0.0507 | 0.0456 | 0.0458 | 0.0491 | 0.0500 | 0.0447 | 0.0449 |
| | (0.0207) | (0.0203) | (0.0203) | (0.0203) | | | | |
| \( \Omega_{1,2} \) | 0.1287 | 0.0951 | 0.1235 | 0.1249 | 0.1445 | 0.0919 | 0.1242 | 0.1266 |
| | (0.2308) | (0.2145) | (0.0241) | (0.2201) | | | | |
| \( \Omega_{1,3} \) | -0.9827 | -0.9825 | -0.9806 | -0.9810 | -0.9999 | -0.9999 | -0.9969 | -0.9970 |
| | (0.0284) | (0.0278) | (0.0278) | (0.0278) | | | | |
| \( \Omega_{2,2} \) | 0.0233 | 0.0243 | 0.0204 | 0.0205 | 0.0231 | 0.0246 | 0.0201 | 0.0204 |
| | (0.0117) | (0.0117) | (0.0117) | (0.0117) | | | | |
| \( \Omega_{2,3} \) | -0.1333 | -0.0963 | -0.1244 | -0.1258 | -0.1428 | -0.0933 | -0.1240 | -0.1258 |
| | (0.2286) | (0.2393) | (0.2393) | (0.2393) | | | | |
| \( \Omega_{3,3} \) | 0.0512 | 0.0517 | 0.0464 | 0.0465 | 0.0493 | 0.0498 | 0.0446 | 0.0448 |
| | (0.0211) | (0.0207) | (0.0207) | (0.0207) | | | | |