TI 2023-073/III
Tinbergen Institute Discussion Paper

Long-term investing under uncertain parameter instability

Bart Keijsers[1]

1 University of Amsterdam and Tinbergen Institute

Long-term investing under uncertain parameter instability*

Bart Keijsers†
University of Amsterdam
Tinbergen Institute

November 2, 2023

Abstract

The relationship between excess returns and the dividend price ratio is known to be unstable. However, there is no consensus on the type of instability, i.e. few or many breaks. Differences in parameter instability affect the long-term investor in particular, as misspecification errors are exacerbated as the investment horizon increases. Therefore, we investigate the consequences of different types of break processes for a long-term investor. The break process is inferred with a mixture innovation model using Bayesian methods. This allows us to estimate the break risk and the uncertainty around it. The estimated parameters show substantial instability, with an average break probability of 20.6%. Assuming constant parameters can lead to losses of up to 16.3% in certainty equivalent return for the long-term investor, even if the break probability is small in reality. The costs of ignoring uncertainty regarding the instability are smaller, but non-negligible.

Keywords: Return predictability, parameter instability, mixture innovation model, long-term investing, Bayesian modeling

JEL classification: C11, C32, G11

*I would like to thank Dick van Dijk, Bart Diris, Graham Elliott, John Maheu, Richard Paap, Davide Pettenuzzo, Francesco Ravazzolo, Allan Timmermann, Rossen Valkanov, Michel van der Wel, participants at the 5th PFMC 2017, the 28th (EC)^2 Amsterdam 2017, the EWMES Barcelona 2017, the 8th ESOBE Maastricht 2017, and seminar participants at Erasmus University Rotterdam, University of Manchester, University of Gothenburg, ECB, and University of Amsterdam for helpful discussions and comments. Part of this research was carried out while visiting the University of California, San Diego. I would like to thank Allan Timmermann for his hospitality. I gratefully acknowledge financial support from the Prins Bernhard Cultuurfonds and the Erasmus Trustfonds.

†b.j.l.keijsers@uva.nl.

1 Introduction

There is substantial evidence for instabilities or time-variation in the relationship between stock returns and potential predictors, see e.g. Pesaran and Timmermann (1995); Rapach and Wohar (2006); Paye and Timmermann (2006); Lettau and Van Nieuwerburgh (2008); Ang and Bekaert (2007). Instabilities are a potential explanation for the lack of sustained predictability (Rapach and Zhou, 2013). The time-variation could be due to changes in economic conditions or learning by economic agents, among others. By accounting for the time-variation, one can again find periods of predictability that can be exploited (Farmer et al., 2023). This is important to investors, since their portfolio depends on the expected risk and return on stocks. Suggested methods that allow for time-varying parameters assume few breaks (Henkel et al., 2011; Pettenuzzo and Timmermann, 2011), or breaks each period (Dangl and Halling, 2012; Johannes et al., 2014). But whether there are few or many breaks implies different properties for returns that are relevant for investors. Should they anticipate a few possibly large shocks, or will the relationship change gradually over time and allow them time to adapt?

Therefore, in the first part of the paper, we aim to find out *type of time-variation* that is supported by the data, few or many breaks. This is achieved by estimating a mixture innovation (MI) model (McCulloch and Tsay, 1993; Giordani and Kohn, 2008), a flexible type of state space models where parameters can change each period, but don't have to. The time-variation is governed by the break probability parameter π, our measure of parameter instability. Based on monthly US data from 1946 to 2015, we find substantial time-variation in the relationship between the dividend price ratio and stock returns. The posterior mode of the break probability is only 2.6%, but the mean is 20.6%. It is hard to clearly identify large breaks, and there is quite some uncertainty regarding the break probability due to the low signal-to-noise ratio. These results are robust to changes in the break probability prior.

In the second part of the paper, we investigate the economic consequences of misspecifying the type of instability in the relationship between the dividend price ratio and excess returns, motivated by Barberis (2000). This is achieved by interpreting the

prior on the break probability as the investor's views regarding the instability. Instability in the relationship between predictors and the excess returns is arguably more important to the long-term investor than for the short-term investor, as the risk of breaks is exacerbated over the investment horizon. Even if the break probability is small, the probability of a break occurring in the period of holding the portfolio can still be high. For example, if the break probability is 5%, the break risk for the one period investor is reasonably small. For an investor holding a portfolio for 120 periods, or ten years, the expected number of breaks in this period is 6. Therefore, we economically evaluate the different views for a long horizon investor with power utility preferences by computed the loss in certainty equivalent return from assuming an incorrect type of time-variation.

We find large costs associated with ignoring a time-varying relationship between the dividend price ratio and stock returns. For a buy-and-hold investor with power utility and risk aversion of $\gamma = 5$, the loss can run up to 16.3% in annualized certainty equivalent return at a 20 year investment horizon. In contrast, the costs from instability misspecification are limited if some type of time-variation is included in the model. A model with many breaks is slightly preferred, likely due to uncertainty in the number and location of breaks. Seemingly in contrast to Johannes et al. (2014), it is more important to take instability in the loadings into account, rather than in the variance. This difference can be explained the fact that we focus on the longer horizon. Also, in line with Pástor and Stambaugh (2012); Pettenuzzo and Timmermann (2011); Johannes et al. (2014), the additional layer of break probability uncertainty implies that the predictive density increases over the investment horizon. This seems in contrast with Carvalho et al. (2018), but in their appendix they also find upward sloping predictive variances for monthly data.

This paper relates to the long literature on return predictability. Welch and Goyal (2008) famously show that popular predictors of stock returns perform poorly out-of-sample. One of the explanations for the poor performance is that the relationship between predictors and the equity premium varies over time. Timmermann (2008) argues that predictive content does exists occasionally. But when it appears, it is quickly exploited by market participants, thus creating short-lived "pockets of predictability"

(Farmer et al., 2023). Multiple authors explicitly model the time-variation. In doing so, they choose either to assume few breaks or many breaks. On the one hand, Ang and Bekaert (2002), Guidolin and Timmermann (2007) and Henkel et al. (2011) estimate a Markov switching model with two to four recurring regimes. Their findings suggest that the predictive power is countercyclical, it is stronger during recessions. Pettenuzzo and Timmermann (2011) consider a change point model, where it is not possible to return to a past regime. On the other hand, Dangl and Halling (2012), Johannes et al. (2014) and Diris (2014) consider a time-varying parameter model where the parameter changes each time period. This typically leads to smoother variations over time. We contribute to this literature by explicitly estimating the type of instability, without assuming the number of breaks ex ante.

Our research adds to the long-term investment literature (Campbell and Viceira, 2002), in particular on the effects of modeling choices and associated uncertainties on portfolios. Kandel and Stambaugh (1996) and Barberis (2000) study the consequences of parameter uncertainty, and Avramov (2002) and Cremers (2002) do this for model uncertainty. Pettenuzzo and Timmermann (2011), Johannes et al. (2014) among others show the impact of parameter instability. We contribute in assessing the impact of uncertain parameter instability, where the type of time-variation is uncertain. Moreover, we compare the economic loss not just with the static case, but also between different types of time-variation.

The methodology is most closely related to Koop et al. (2009), who, in a monetary policy application, estimate a MI model with a similar break process specification. They share our motivation of trying to let the data speak on the parameter instability. We apply it in a financial setting with low signal-to-noise ratio, and include persistence in the time-varying parameters.

Finally, in spirit, the paper is related to Clark and Ravazzolo (2015), Bauwens et al. (2015) and Pettenuzzo and Timmermann (2017) in comparing different types of time-varying models instead of just comparing to a static model. These papers statistically compare (mostly univariate) macroeconomic models. This work differs in

that we economically evaluate the performance in a multivariate model for financial time series, with a low signal-to-noise ratio.

The paper is structured as follows. Section 2 defines the mixture innovation model, details the priors and estimation procedure. Section 3 briefly describes the data. Section 4 presents the estimation results, followed by the economic evaluation in Section 5. Section 6 is a sensitivity analysis on the break probability prior, and Section 7 concludes.

2 Methodology

In line with the literature on return predictability (Rapach and Zhou, 2013) and portfolio choice (Brandt, 2010), the relationship between the excess log return r_t and the predictor z_t is described by a restricted vector autoregressive (VAR) model,

$$r_t = \beta_{1t} + \beta_{2t} z_{t-1} + \varepsilon_{1t}, \tag{1}$$

$$z_t = \beta_{3t} + \beta_{4t} z_{t-1} + \varepsilon_{2t}, \tag{2}$$

for $t = 1, \ldots, T$, with T the sample size, where $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \varepsilon_{2t})' \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Omega}_t)$. The intercepts and loadings $\boldsymbol{\beta}_t = (\beta_{1t}, \ldots, \beta_{4t})'$ and covariance matrix $\boldsymbol{\Omega}_t$ are potentially time-varying. Following Primiceri (2005), among others, the covariance matrix $\boldsymbol{\Omega}_t$ is decomposed as $\boldsymbol{A}_t \boldsymbol{\Omega}_t \boldsymbol{A}_t' = \boldsymbol{\Sigma}_t \boldsymbol{\Sigma}_t'$, where $\boldsymbol{A}_t = \begin{bmatrix} 1 & 0 \\ \alpha_t & 1 \end{bmatrix}$ is a lower triangular matrix with covariance term α_t, and $\boldsymbol{\Sigma}_t = \mathrm{diag}(\sigma_{1t}, \sigma_{2t})$ a diagonal matrix with the (structural) volatilities σ_{1t} and σ_{2t} on the diagonal. This decomposition simplifies inference, because α_t is not restricted between zero and one. It also allows us to rewrite the model in (conditionally) Gaussian state space form to easily draw the variance and covariance terms.

2.1 Modeling parameter instability

To describe the time-variation in the parameters, we employ a mixture innovation (MI) model, introduced by McCulloch and Tsay (1993). It is a flexible conditionally Gaussian

state space model, where each time-varying parameter θ_t is modeled as,

$$\theta_t = \theta_{t-1} + \kappa_t \eta_t, \tag{3}$$

where $\eta_t \sim \mathrm{N}(0, \sigma_\eta)$, such that the break size depends on the break process κ_t. Equation (3) excludes a persistence parameter for illustrative purposes. We focus on a binary break process, $\kappa_t \in \{0, 1\}$, which is i.i.d. with break probability $\Pr[\kappa_t = 1] = \pi$.

To see how the process in Equation (3) works, consider what happens for different values of the break process κ_t. If $\kappa_t = 0$, then $\theta_t = \theta_{t-1}$, so there is no break and the parameter stays in the current regime. If $\kappa_t = 1$, then $\theta_t = \theta_{t-1} + \eta_t$, there is a break and we are in a new regime. This break process provides flexibility and differentiates it from alternative models. The break probability π and break size σ_η define the type of instability.

We use an MI model for our analysis because of its attractive properties. First and most importantly, the break probability π represents the parameter instability – our measure of interest. Moreover, it is intuitive to interpret, which facilitates prior elicitation. Second, the MI model nests many types of time-variation, such as few (large) breaks or many (small) breaks. As extreme cases, the constant parameter model when $\pi \to 0$, and the time-varying parameter (TVP-VAR) model (Cogley and Sargent, 2001, 2005; Primiceri, 2005) when $\pi \to 1$ are included. The MI model does not nest the classic change point model (Chib, 1998), because the number of breaks is not fixed. But the version with an unknown number of breaks (Koop and Potter, 2007) is approximated by a small value for π. Hence, this provides a framework to easily compare these models for the long-term investor.

Third, there is no need to select the type of time-variation, or the number of breaks, ex ante. Estimating the break probability, we infer the type of instability supported by the data and the posterior variance provides the uncertainty regarding this instability. Fourth, independent break processes for different parameters can be specified. And indeed, as we will explain below, the parameters in the mean, variance and covariance are governed by different break processes.

Fifth, inference is computationally efficient due to the algorithm of Gerlach et al. (2000), which is applicable to any conditionally Gaussian state space model. In state space terminology, Equations (1) and (2) are the observation or measurement equations (Durbin and Koopman, 2012). The efficiency of Gerlach et al.'s (2000) algorithm comes from drawing κ_t without conditioning on the states θ_t. Drawing the break process κ_t is linear in the number of values the break process can take. It requires $O(T2^K)$ operations, with T the sample size and K the number of binary break processes (Giordani and Kohn, 2008). In some cases can even be improved further (Giordani and Kohn, 2008; Fiorentini et al., 2014). In contrast, a change point model with an unknown number of breaks requires $O(T^2)$ operations, with $T > K$ (Koop and Potter, 2007).

2.2 Mixture innovation model

To adapt the model in Equations (1)–(2) to a MI model, observe that the parameters can naturally be divided into three groups, that may be subject to different types of time-variation: (i) intercepts and loadings, (ii) residual variances, and (iii) residual covariance. To capture the grouping, we implement a mixture innovation model with three independent binary break processes, one for each parameter type. Koop et al. (2009) use a similar specification in a macroeconomic application.

The intercepts and loadings $\boldsymbol{\beta}_t$ follow a mixture innovation process that share one break process. To model persistence of $\boldsymbol{\beta}_t$, a modification to Equation (3) is required. The state equation for the intercepts and loadings $\boldsymbol{\beta}_t$ is

$$\boldsymbol{\beta}_t = \boldsymbol{\mu}_b + f(\kappa_{bt}, \boldsymbol{\Phi}_b)(\boldsymbol{\beta}_{t-1} - \boldsymbol{\mu}_b) + \kappa_{bt}\boldsymbol{\eta}_t, \tag{4}$$

for $t = 2, \ldots, T$, where $\boldsymbol{\eta}_t = (\eta_{1t}, \ldots, \eta_{4t})' \sim \mathrm{N}(\mathbf{0}, \boldsymbol{Q}_b)$ with break size matrix $\boldsymbol{Q}_b$, κ_{bt} is a binary break process which is one if there is a break at time t and zero otherwise that is i.i.d. with break probability $\Pr[\kappa_{bt} = 1] = \pi_b$, $\boldsymbol{\mu}_b = (\mu_{b,1}, \ldots, \mu_{b,4})'$ is a vector of long run

means, and $f(\kappa_{bt}, \boldsymbol{\Phi}_b)$ is an autoregressive (AR) function,

$$f(\kappa_{bt}, \boldsymbol{\Phi}_b) = \begin{cases} \boldsymbol{I}_{n_k} & \text{if } \kappa_{bt} = 0 \\ \boldsymbol{\Phi}_b & \text{if } \kappa_{bt} = 1, \end{cases} \tag{5}$$

where $\boldsymbol{I}_{n_k}$ is the $n_k \times n_k$ identity matrix, with $n_k = 4$ the parameter vector length, and $\boldsymbol{\Phi}_b = \text{diag}(\phi_{b,1}, \ldots, \phi_{b,4})$ a diagonal matrix with autoregressive parameters. We refrain from estimating the off-diagonal elements, to limit the number of parameters. The covariance $\boldsymbol{Q}_b$ can still be a full matrix.

The outcome of the function in Equation (5) is either the identity matrix if there is no break, or a diagonal matrix with autoregressive parameters if there is a break at time t. Hence, if there is no break we have the same value as in the previous period, but if there is a break, we have an autoregressive process. The autoregressive MI (ARMI) process in Equation (4) nests the AR process as a special case, if $\pi_b = 1$, and the MI process in Equation (3) if $\boldsymbol{\Phi}_b = \boldsymbol{I}_{n_k}$. In case $\boldsymbol{\Phi}_b$ is diagonal, the process is stationary if $|\phi_i| < 1$, for all $i = 1, \ldots, n_k$, see Appendix A.

The autoregressive function is included to model the parameters' persistence. Many applications of the MI model (or TVP model) assume a random walk by setting $f(\kappa_{bt}, \boldsymbol{\Phi}_b) = \boldsymbol{I}$, to reduce the number of parameters, see e.g. Primiceri (2005), Koop et al. (2009) or Groen et al. (2013). In that case though, the parameter variance and hence also the predictive density's variance grows linearly with the forecast horizon and is unbounded in the limit. This is undesirable in our setting, because the long-term investor is concerned with the return density many periods into the future. Therefore, the prior on $\boldsymbol{\Phi}$ excludes the random walk and explosive processes, see Appendix Section B.

The state equations for the (log) variances $\boldsymbol{\sigma}_t^2 = (\sigma_{1t}^2, \sigma_{2t}^2)'$ and the covariance term α_t are

$$\log \boldsymbol{\sigma}_t^2 = \boldsymbol{\mu}_s + f(\kappa_{st}, \boldsymbol{\Phi}_s)(\log \boldsymbol{\sigma}_{t-1}^2 - \boldsymbol{\mu}_s) + \kappa_{st}\boldsymbol{\zeta}_t, \tag{6}$$

$$\alpha_t = \mu_a + f(\kappa_{at}, \phi_a)(\alpha_{t-1} - \mu_a) + \kappa_{at}\xi_t, \tag{7}$$

for $t = 2, \ldots, T$, where $\boldsymbol{\zeta}_t = (\zeta_{1t}, \zeta_{2t})' \sim \mathrm{N}(\mathbf{0}, \boldsymbol{Q}_s)$ and $\xi_t \sim \mathrm{N}(0, q_a^2)$ independent from $\boldsymbol{\varepsilon}_t$, $\boldsymbol{\eta}_t$ and each other, κ_{st} and κ_{at} are binary break processes that are i.i.d. with break probabilities $\Pr[\kappa_{st} = 1] = \pi_s$ and $\Pr[\kappa_{at} = 1] = \pi_a$, $\boldsymbol{\mu}_s = (\mu_{s,1}, \mu_{s,2})'$ and μ_a are the long run means, and $f(\cdot, \cdot)$ the autoregressive function as defined in Equation (5), with $\boldsymbol{\Phi}_s = \mathrm{diag}(\phi_{s,1}, \phi_{s,2})$ and ϕ_a the autoregressive parameters. The variance terms are specified in logs to ensure positivity (Kim et al., 1998).

The break processes κ_{bt}, κ_{st} and κ_{at} are independent and allowed to, but not restricted to, break at different points in time. It can be extended to allow for dependence between the breaks, or to allow each parameter to have its own break process. We refrain from this for a few reasons. First, the coefficients $\boldsymbol{\beta}_t$ parameters are quite strongly correlated, so it seems reasonable to assume that they share the same break process. Second, the model likely benefits from some structure given the low signal-to-noise ratio of the data. It therefore doesn't seem to be worth the additional computational costs. This setup does capture potential differences in time-variation between the first and second moment, which allows to specify relevant alternatives such as a model with constant predictability but with stochastic volatility.

2.3 Priors

We employ Bayesian methodology to estimate the MI model in Equations (1)–(2) and Equations (4)–(7). This is a natural way to include parameter uncertainty (Barberis, 2000). Moreover, it allows us to assess the degree of uncertainty on the break probability. Given the Bayesian methodology, we need to specify prior distributions for the static parameters and initial conditions for the state equations. The priors are mostly standard conjugate empirically Bayesian priors, based on Primiceri (2005) and Koop et al. (2009). The location and scaling are based on ordinary least squares estimates of the restricted VAR in Equations (1)–(2) using the first ten years of data, prior to the estimation sample. The priors regarding the initial values of the state variables are fairly informative. The priors are less informative for the long run mean and break probability. The full prior specification is in Appendix B, we focus here on the parameter of interest.

The prior for the break probability π_k is a beta distribution,

$$\pi_k \sim \text{Beta}(\underline{a}_k, \underline{b}_k), \tag{8}$$

for $k \in \{b, s, a\}$. The hyperparameters $\underline{a}_k$ and $\underline{b}_k$ reflect the investor's views on the presence of breaks, or lack thereof.

For the intercepts and loadings, we set the hyperparameters $\underline{a}_k$ and $\underline{b}_k$ such that the prior's expectation is between a few breaks (about one break every ten years) and many breaks (about one break per year). In particular, we set $\underline{a}_b = 1$ and $\underline{b}_b = 59$, which implies an average duration of 60 months between breaks, and a fairly weak prior. The hyperparameters can be thought of as adding $\underline{a} + \underline{b}$ hypothetical observations: $\underline{a}$ ones and $\underline{b}$ zeros, i.e. $\underline{a}$ breaks. If we define the prior strength as $(\underline{a} + \underline{b})/T$, with T the sample size, the prior strength is 7.1%. For comparison, Ravazzolo et al. (2008) and Groen et al. (2013) employ similar priors with prior strengths of 7.4% and 5.2%, respectively.

For the variances and covariance's break processes, we set $\underline{a}_s = \underline{b}_s = \underline{a}_a = \underline{b}_a = 1$. This is an uninformative prior, but effectively puts most probability mass at the many breaks case to favor stochastic volatility.

2.4 Inference

Inference involves a Gibbs sampler. The algorithm mostly follows that of Koop et al. (2009), which is based on Primiceri (2005) with the addition of drawing the break processes $\boldsymbol{\kappa}_t$ and probabilities $\boldsymbol{\pi}$. We add a step to draw the autoregressive parameters $\boldsymbol{\Phi}_b$, $\boldsymbol{\Phi}_s$ and ϕ_a.

The algorithm works as follows. For each of the state variables, $\boldsymbol{\beta}_t$, $\boldsymbol{\sigma}_t$, and α_t, we rewrite the VAR such that we can apply methods for (conditionally) linear Gaussian state space models. Then, we first draw the break process $\kappa_{k,t}$ using the algorithm by Gerlach et al. (2000). Next, we draw the latent variable using the simulation smoother of Durbin and Koopman (2002). After drawing the time-varying parameters, we draw the static parameters (π_k, $\boldsymbol{Q}_k$ and $\boldsymbol{\Phi}_k$, for $k \in \{b, s, a\}$) from their full conditional posterior

distributions. For a detailed description, see Appendix C.

The sampler is iterated until we have 10,000 retained draws, after removing a burn-in sample of 2,000 draws. Increasing the number of draws does not affect results. See also a convergence analysis in Appendix D.

3 Data

We use the monthly S&P 500 index return and the one-month T-bill rate as risk-free rate to construct the excess log return. These are taken from the extended Welch and Goyal (2008) data set, available on Amit Goyal's website: `http://www.hec.unil.ch/agoyal/`. The sample consists of monthly observations from January 1936 to December 2021. It is split into two parts. The first ten years, January 1936 to December 1945, are used for prior calibration. The second part is the estimation sample of 912 months.

As predictor, we use the (log) dividend price ratio, defined as the ratio of the sum of the dividends over the last 12 months to the current stock price. It is one of the most popular predictors in the literature, has shown to hold predictive power (Campbell and Shiller, 1988), and has been used in most studies focusing on parameter instability. Summary statistics and a time series plot are in Appendix E. The time series suggests a level shift in the 1990s for the dividend price ratio (cf. Lettau and Van Nieuwerburgh, 2008). This break could disrupt the relationship with the equity premium, which the MI model can capture by allowing for instability in the intercepts and loadings in Equations (1)–(2).

4 Estimation results

4.1 Few versus many breaks

To assess the impact of different break probabilities, we first estimate the mixture innovation model with a known break probability. That is, we assume various values for a fixed break probability π_b, from no to few to many breaks, $\pi_b \in \{0, 1/600, 1/120, 1/60, 1\}$. This includes the extreme cases of a static, i.e. no breaks ($\pi_b = 0$), and a TVP-VAR

($\pi_b = 1$) model. The few breaks cases are motivated by previous studies: $\pi_b = 1/120$ (one break expected every 10 years) approximates the 8 to 10 breaks estimated by Pettenuzzo and Timmermann (2011), and $\pi_b = 1/600$ (one break expected every 50 years) is in line with the one to two breaks found by Lettau and Van Nieuwerburgh (2008). Fixing the break probability at $\pi_b = 1/60$ (one break expected every 5 years) is an intermediate case between the few breaks and the many breaks. The break probability for the variances and covariance term is set at $\pi_s = \pi_a = 1$, to capture the volatility clustering. Only in the case of $\pi_b = 0$, the model is fully static, ignoring parameter instability, so $\pi_k = 0$, for $k \in \{b, s, a\}$.

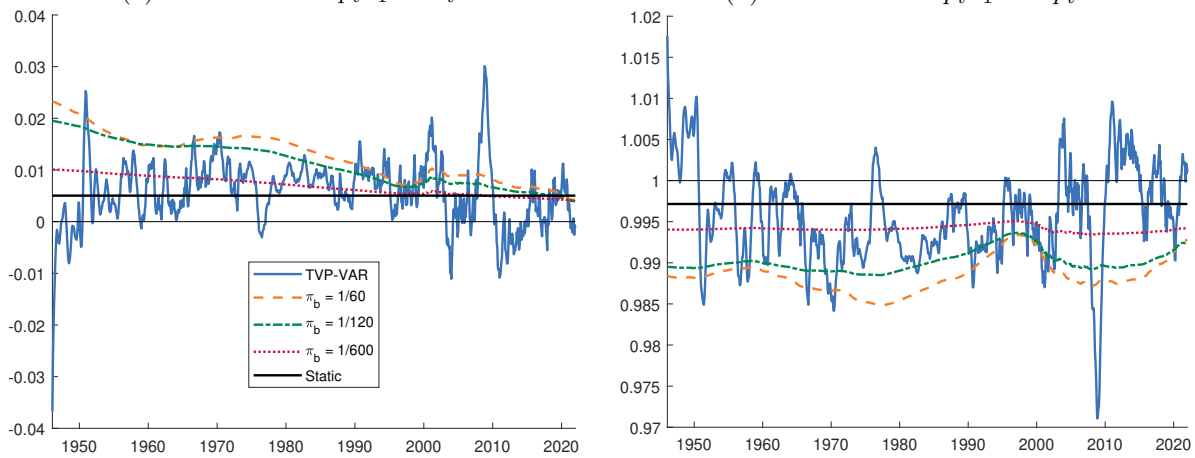
The posterior estimates of the time-varying parameters in Figures 1 and 2 are in line with previous research. First, Figure 1a confirms that the weak but positive predictive power of the dividend price ratio has decreased from the 1970s, and in particular small in the 1990s (see e.g. Ang and Bekaert, 2007). Second, the static model estimates the predictability coefficient in Figure 1a to be smaller compared to models that do allow for time-varying parameters. The conditional level of predictive power is probably underestimated by the static model, as it is a smoothed estimate over a sample with periods where predictive power fluctuates. Including the periods without predictive power shrink the estimate downwards. Third, the estimates illustrate why the dividend price ratio is often used in long-term investing literature. The high persistence (see Figure 1b), combined with positive predictability coefficient in Figure 1a and strong negative residual correlation in Figure 2b implies strong mean reversion, making stocks safer in the long run and ensuring a sizable hedge term. Fourth, Figure 2a shows that volatility clustering is captured by all models, outside the static model. We recognize periods of high volatility, such as the Oil Crisis in the 70s, Black Monday on October 1987, the financial crisis in 2008, and the recent COVID-19 period.

Comparing across the break probabilities, the results in Figure 1 are in line with the interpretation of π_b. As the break probability increases, the posterior estimate displays more instability. The TVP-VAR's estimate is much more volatile than for any of the other models, and may be fitting noise rather than signal. The estimates for the other

types of time-variation are largely similar to each other. If we expect one or two breaks ($\pi_b = 1/600$), the intercepts and loadings are quite stable. Interestingly, Figure 1b provides some evidence of a break in the dividend price ratio at the end of the 1990s or start of 2000s, conform Lettau and Van Nieuwerburgh (2008). The posterior mean is quite smooth when $\pi = 1/120$ and $\pi = 1/60$, indicating uncertainty regarding the break locations.

Figure 1: Posterior mean of intercepts and loadings under known π

(a) Coefficient of dp_{t-1} on r_t (b) Coefficient of dp_{t-1} on dp_t

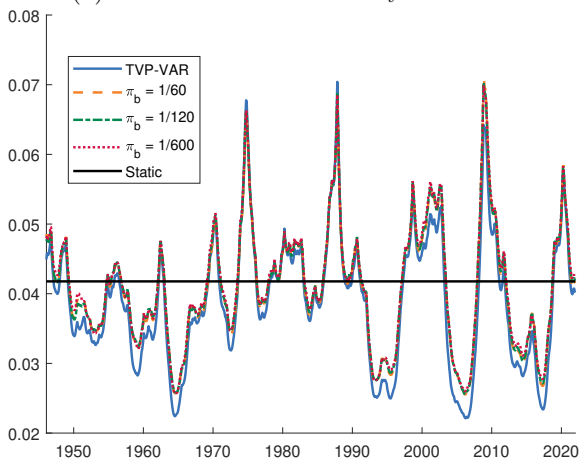


The figures present the posterior mean of the time-varying loadings, β_{2t} and β_{4t}, from the mixture innovation (MI) model in Equations (1)–(2) with the time-varying parameters described by Equations (4)–(7), with break probability for intercepts and loadings fixed at $\pi_b \in \{0, 1/600, 1/120, 1/60, 1\}$. The break probability for the variances and the covariance term is fixed at $\pi_s = \pi_a = 1$ for all models, except for the static case ($\pi_b = 0$), where $\pi_s = \pi_a = 0$. TVP-VAR is the model with $\pi_b = \pi_s = \pi_a = 1$. See Section 2.3 for the prior specifications.

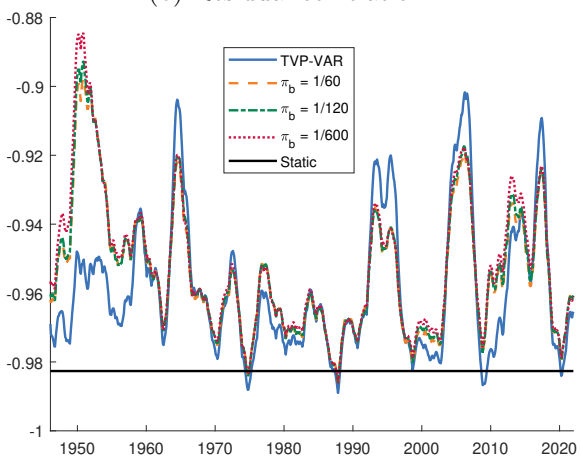
To illustrate that the different break probabilities imply different statistical properties for the returns, Figures 3a–3b show individual draws for the MI model when $\pi = 120$ and the TVP-VAR model. Draws from the TVP-VAR model are very noisy, with a break each period. In contrast, the draws under a small break probability exhibit large periods of no change, with relatively large breaks. Some draws display noisy periods near the end of the 1990s, perhaps because of uncertainty regarding a break around that time. These draws are representative of the individual draws that feed into the predictive density. Their large differences motivate us to assess the consequences in terms of asset allocation in Section 5.

Figure 2: Posterior mean of residual variance and correlation under known π

(a) Standard deviation of r_t residuals



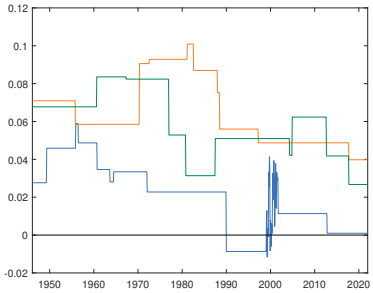
(b) Residual correlation



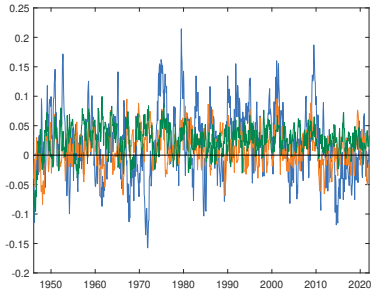
The figures present the posterior mean of the residual variance of r_t, σ_{1t}, and correlation from the mixture innovation (MI) model in Equations (1)–(2) with the time-varying parameters described by Equations (4)–(7), with break probability for intercepts and loadings fixed at $\pi_b \in \{0, 1/600, 1/120, 1/60, 1\}$. The break probability for the variances and the covariance term is fixed at $\pi_s = \pi_a = 1$ for all models, except for the static case ($\pi_b = 0$), where $\pi_s = \pi_a = 0$. TVP-VAR is the model with $\pi_b = \pi_s = \pi_a = 1$. See Section 2.3 for the prior specifications.

Figure 3: Three posterior draws of predictability coefficient

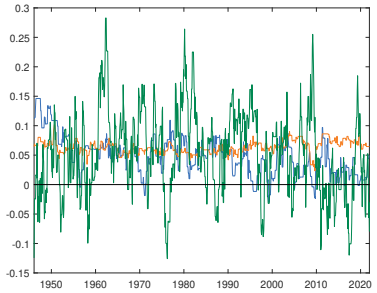
(a) $\pi_b = 1/120$



(b) TVP-VAR



(c) π unknown



The figures present three random posterior draws of β_{2t}, the coefficient of dp_{t-1} on r_t, for (a) the mixture innovation model, with break probability fixed at $\pi_b = 1/120$, (b) the TVP-VAR model, and (c) the MI model with unknown break probability. See Section 2.3 for the prior specification.

4.2 Estimating the parameter instability

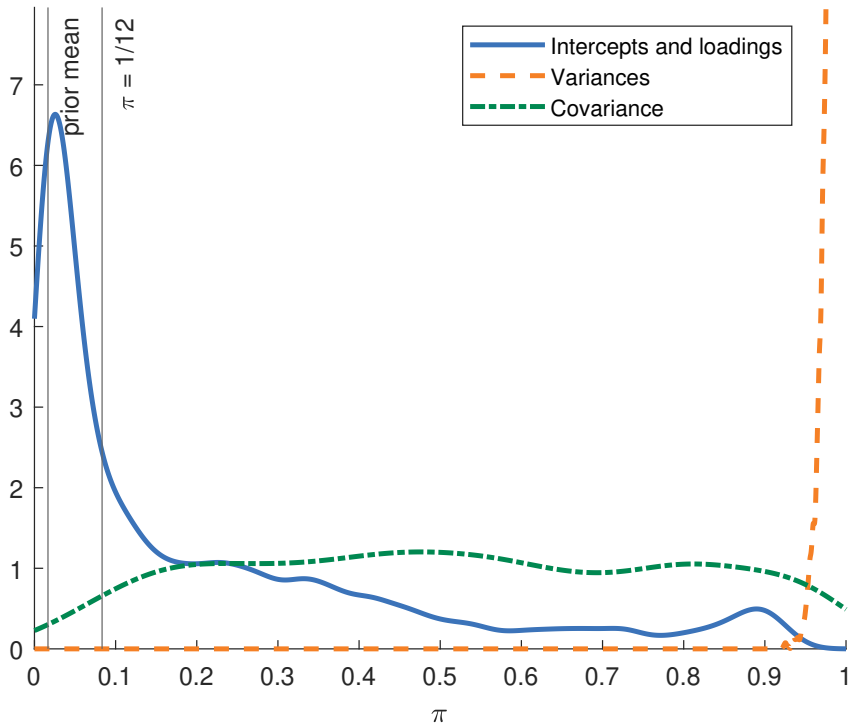
The break probability is unknown in practice, emphasized by the lack of consensus in the literature. Therefore, we specify the prior for $\boldsymbol{\pi}$ as in Section 2.3, and let the data speak about the type of instability, and the uncertainty around it.

Figure 4 presents the posterior of the break probability. Most probability mass is near $\pi_b = 0$. The mode is at 2.6%, smaller than 5%. The posterior mean of the break probability of the coefficients' break process is 20.6%, which implies an expected duration between breaks of about 5 months. The expected prior break probability is 1.7%, so the data suggests a fairly large number of breaks. At the same time, the density is wide,

with a standard deviation of 21.1%. It is substantially larger than the prior standard deviation of 1.6%. This indicates quite some uncertainty regarding the break probability.

It is good to keep in mind that due to the parameterization of the MI model, there is relatively much support of π associated with a large number of breaks. Even in the prior with expected break probability $\pi_b = 1/600$ there is a non-negligible 2.4% probability that $\pi_b > 1/60$, i.e. that there are is more than one break every five years. This perhaps makes it easier for the posterior to be pushed to the many breaks case. In contrast, with a change point model it would be easier to allocate zero probability mass to a large number of breaks. The change point model does not yield a direct measure of instability though, nor with a coherent framework to compare different types of time-variation. Also, computational issues arise when estimating a change point model with many breaks.

Figure 4: Posterior of break probability



The figure presents the posterior densities of the break probabilities from the mixture innovation model in Equations (1)–(2) with the time-varying parameters described by Equations (4)–(7), where the break probability is assumed to be unknown. The hyperparameters for the break probability of intercepts and loadings set at $\underline{a}_b = 1$ and $\underline{b}_b = 59$, and for the break probability of the variances and the covariance term set at $\underline{a}_s = \underline{b}_s = \underline{a}_a = \underline{b}_a = 1$. See Appendix B for the other prior specifications.

There is substantial time-variation in the loadings according to the MI model, more than implied by the few breaks case ($\pi_b = 1/120$ or $\pi_b = 1/600$), see Figure 1. On the other hand, it is more stable than in the TVP-VAR model. The pattern of β_{2t} is similar

to the other MI models, which is mildly countercyclical. E.g. Dangl and Halling (2012) also find a countercyclical predictive relationship. The correlation between the posterior mean and the NBER recession dummy is 14.1% for the full sample, and increases to 21.8% in the period 1959−2021. There is also a sizable correlation of 53.6% with the National Financial Conditions Index (NFCI), for the period when the NFCI is available, from 1971.

Although the posterior mean seems like a smoother version of the TVP-VAR model's, the individual draws in Figure 3c highlight the difference between the models. Some draws resemble those of the TVP-VAR model, whereas other draws display periods of stability, in line with draws from a model with a small break probability. The variety between the draws is a reflection of the wide posterior break probability distribution in Figure 4. The combination of only few large breaks and a wide posterior distribution for the break probability suggests that it is hard to identify the break probability with high certainty, probably due to the low signal-to-noise ratio. This highlights the difficulty in predicting stock returns.

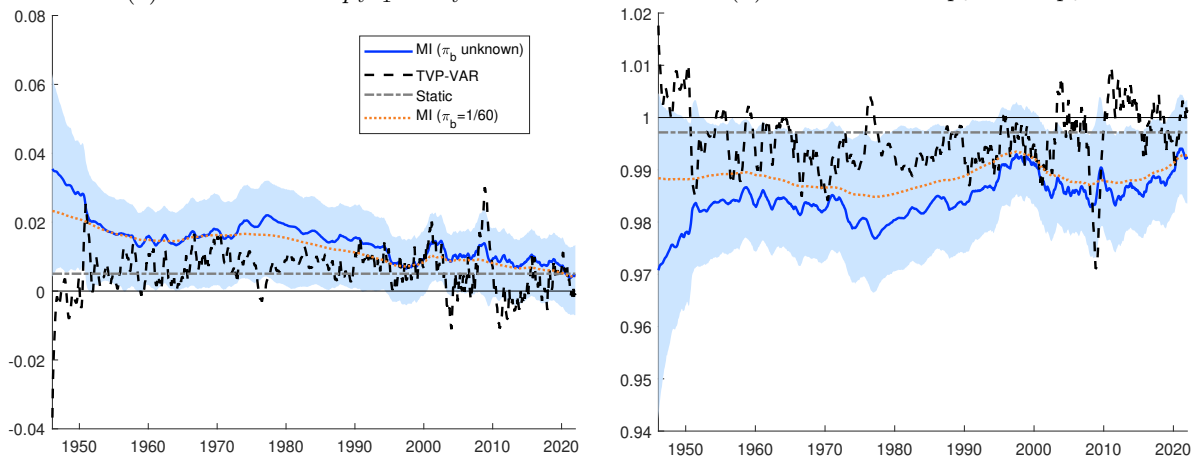
The variance for the MI model with unknown break probability in Figure 6 is essentially the same as for the TVP-VAR model and for the MI model with fixed break probability. The MI model confirms the presence of heteroskedasticity, as the posterior of the break probability in Figure 4 strongly suggests stochastic volatility rather than a few regime switches. Johannes et al. (2014) find that including stochastic volatility is needed to improve the predictive ability.

4.3 In-sample fit

We compare the fit of the models using the Watanabe Information Criterion (WAIC, Watanabe, 2010), a Bayesian measure of fit that penalizes the number of parameters. The estimated effective number of parameters varies depending on the time-variation in the model. It is computed as the variance of the observation level loglikelihood over the posterior draws, summed over all observations. This is known as WAIC type 2, which is quite stable and has the attractive property of being related to leave-one-out

Figure 5: Posterior mean of intercepts and loadings under unknown π

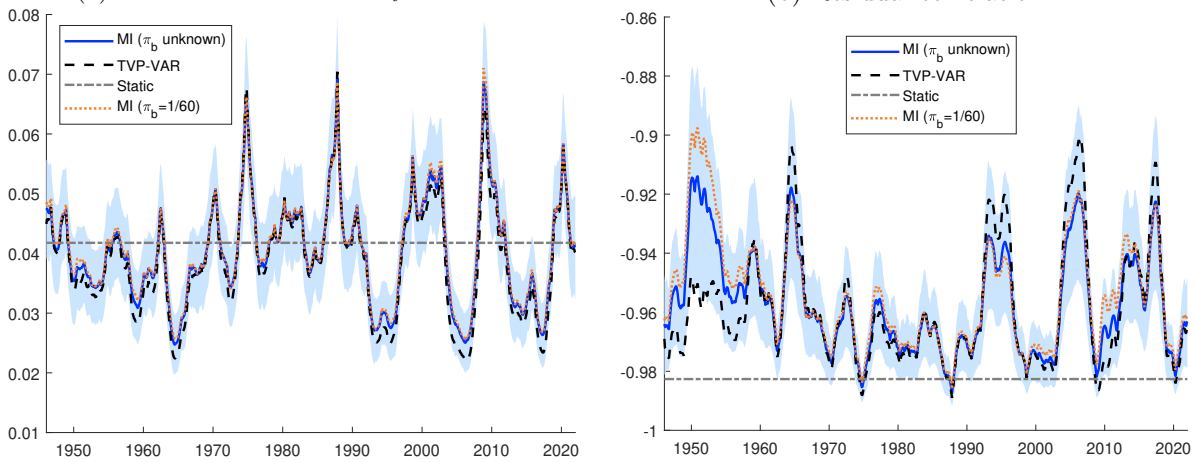
(a) Coefficient of dp_{t-1} on r_t (b) Coefficient of dp_{t-1} on dp_t



The figures present the posterior mean of the time-varying loadings, β_{2t} and β_{4t}, from the mixture innovation (MI) model in Equations (1)–(2) with the time-varying parameters described by Equations (4)–(7). If π_b is assumed unknown (solid blue line), the hyperparameters for the break probability of intercepts and loadings set at $\underline{a}_b = 1$ and $\underline{b}_b = 59$, and for the break probability of the variances and the covariance term set at $\underline{a}_s = \underline{b}_s = \underline{a}_a = \underline{b}_a = 1$. If the break probability is assumed known (orange dashed line), it is fixed at the prior mean, such that $\pi_b = 1/60$ and $\pi_s = \pi_a = 1$. TVP-VAR (dashed black line) is the model with $\pi_b = \pi_s = \pi_a = 1$, and static (dashed-dotted gray line) is the model where $\pi_b = \pi_s = \pi_a = 0$. The blue shaded area is the 68% credible interval for the MI model where π_b is unknown. See Appendix B for the other prior specifications.

Figure 6: Posterior mean of residual variance and correlation under unknown π

(a) Standard deviation of r_t residuals (b) Residual correlation



The figures present the posterior mean of the residual variance of r_t, σ_{1t}, and correlation from the mixture innovation (MI) model in Equations (1)–(2) with the time-varying parameters described by Equations (4)–(7). If π_b is assumed unknown (solid blue line), the hyperparameters for the break probability of intercepts and loadings set at $\underline{a}_b = 1$ and $\underline{b}_b = 59$, and for the break probability of the variances and the covariance term set at $\underline{a}_s = \underline{b}_s = \underline{a}_a = \underline{b}_a = 1$. If the break probability is assumed known (orange dashed line), it is fixed at the prior mean, such that $\pi_b = 1/60$ and $\pi_s = \pi_a = 1$. TVP-VAR (dashed black line) is the model with $\pi_b = \pi_s = \pi_a = 1$, and static (dashed-dotted gray line) is the model where $\pi_b = \pi_s = \pi_a = 0$. The blue shaded area is the 68% credible interval for the MI model where π_b is unknown. See Appendix B for the other prior specifications.

cross-validation (Gelman et al., 2014). A smaller WAIC value indicates a better model fit.

The WAIC values in Table 1 shows that the static model performs worst, followed by the MI-b model. Both models restrict the parameters $\boldsymbol{\beta}_t$ to be constant. They differ in that MI-b includes time-varying volatility. The TVP-VAR model provides the best fit. The MI model where the break probability π is estimated is a close second. Somewhat surprising, the estimated effective number of parameters is larger than for the TVP-VAR model. This may be due to the difference in break probability over the draws. Thus, the WAIC values provide in-sample evidence for parameter instability. They point towards a more flexible model with many breaks as a more accurate description of the data, even when taking into account the additional parameters needed to estimate.

As a way to assess the model fit for the volatility, we compare the estimated variance with the realized variance (RV). The monthly RV is computed as the sum of the squared daily returns, given by the CRSP value weighted index return, incl. dividends, from 1954 to 2021. The variances are compared in logs.The mean squared error (MSE) in the last column of Table 1 confirms that a more flexible model provides a better fit. The static model performs worst with an MSE of 1.13, while it ranges from 0.68 (TVP-VAR) to 0.75 (MI-b) for models that do allow for time-varying volatility. For reference, the MSE of an AR model for the log realized variance is 0.40. The difference is that this simple AR model uses the realized measure as input, while the MI models only use monthly data. As Figure 7 shows that the estimate from the MI model slightly overestimates the RV, especially in the first half of the sample. However, this is in line with a moving average of the squared *monthly* returns. So the deviation from RV indeed results from using monthly data.

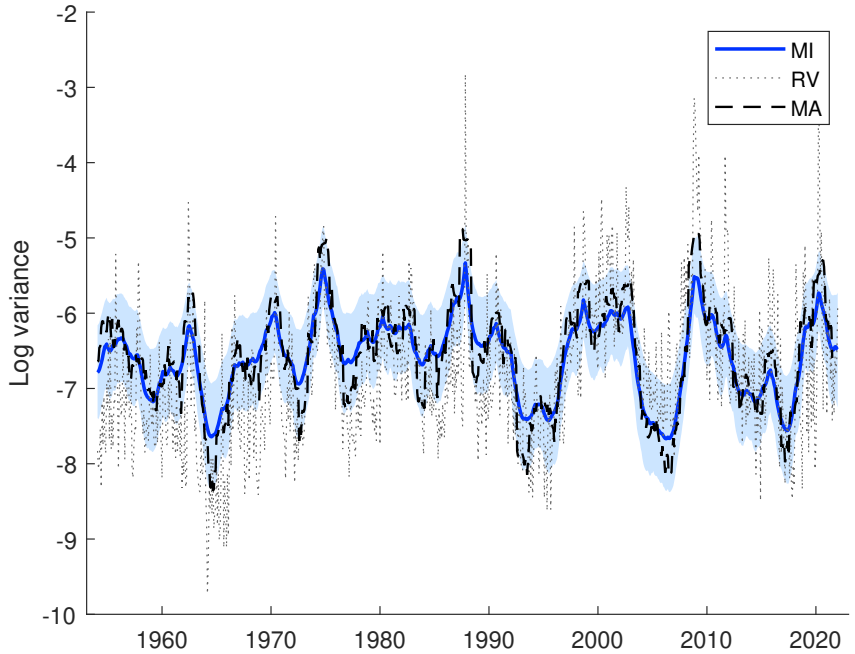
Overall, the results show the importance of allowing for time-variation in both the mean and the variance. That the TVP-VAR model is preferred suggests that either there are indeed many breaks in the relationship between the dividend price ratio and stock returns, or it is a good way to approximate few breaks for which the location is hard to pin down.

Table 1: In-sample fit

Model	Loglikelihood (mean)	$\hat{m}$	WAIC	MSE (log RV)
MI (π unknown)	4,765.7	72.6	−9,450.9	0.718
TVP-VAR	4,783.7	52.7	−9,508.2	0.679
MI ($\pi_b = 1/60$)	4,739.8	67.5	−9,405.1	0.729
MI ($\pi_b = 1/120$)	4,728.8	67.0	−9,383.8	0.741
MI ($\pi_b = 1/600$)	4,709.1	64.7	−9,346.4	0.747
MI-b ($\pi_b = 0$)	4,702.4	55.1	−9,343.7	0.751
Static	4,682.9	5.4	−9,360.4	1.133

The table presents the loglikelihood mean over the posterior draws, the effective number of parameters ($\hat{m}$), and the Watanabe-Akaike Information Criterion (WAIC) for the MI model presented in Section 2 with different prior break probabilities. MI is the mixture innovation model, TVP-VAR is the time-varying parameter VAR model, static is no time-variation and the addition of -b means that the intercepts and loadings are restricted to be constant. The last column presents the mean squared error for the monthly log realized volatility over the period 1955–2021.

Figure 7: Log realized variance fit



The figure presents (in logs) the monthly realized variance (RV), the two-sided 1-year moving average of monthly squared returns (MA), and posterior mean and 90% credible interval of the excess return variance from the mixture innovation model (MI) in Equations (1)–(2) with the time-varying parameters described by Equations (4)–(7), where the break probability is assumed to be unknown. The hyperparameters for the break probability of intercepts and loadings set at $\underline{a}_b = 1$ and $\underline{b}_b = 59$, and for the break probability of the variances and the covariance term set at $\underline{a}_s = \underline{b}_s = \underline{a}_a = \underline{b}_a = 1$. See Appendix B for the other prior specifications.

5 Economic evaluation

In the spirit of Barberis (2000), we analyze the influence of the break probability on the portfolio allocation of a long-term investor. This provides insight into whether the statistical differences in the previous sections translate to economic differences in terms of portfolio risk and returns.

5.1 Asset allocation problem

We consider a power utility (constant relative risk aversion) maximizing buy-and-hold investor with an investment horizon h up to 240 months, or 20 years. So the investor allocates her wealth now, and receives the return at the end of the investment period, without rebalancing. She faces the following problem,

$$\max_{w} U(W_{T+h}) = \max_{w} \frac{W_{T+h}^{1-\gamma} - 1}{1-\gamma}, \tag{9}$$

with risk aversion $\gamma > 0$ and wealth at time $T + h$ is defined as

$$W_{T+h} = W_T \left(w \exp\left(h r_f + \sum_{t=T+1}^{T+h} r_t \right) + (1-w) \exp(h r_f) \right), \tag{10}$$

where w is the fraction of wealth allocated to stocks and r_f is the log risk-free rate, set equal to the one-month T-bill rate historical average. Wealth at time T is normalized to one. The investor optimizes utility by changing the allocation to stocks w, with short selling constraints, such that $0 \leq w \leq 0.99$. The weight $w = 1$ is excluded to avoid unbounded utility, see e.g. Barberis (2000).

To solve the optimization problem in Equation (9), the investor needs to specify the distribution of excess stock returns h periods ahead. We use draws from the posterior to sample from the predictive density of stock returns. Hence, the predictive density incorporates estimation uncertainty, parameter instability, and uncertainty regarding the risk of breaks. Excess returns are simulated up to the investment horizon h, conditional on the model, a time period, and the posterior draws from the m-th MCMC iteration. This is repeated ten times per posterior draw, to increase precision on the moments of the predictive density (Rao-Blackwellization). Using these draws, the expected utility is computed for a given weight w. The optimal weight is then found by a one dimensional adaptive grid search over w, where the step size decreases to a minimum of 10^{-5}.

The maximum horizon h is 240 months, or 20 years, so we want to make sure that the predictive density is stationary. Otherwise, the variance is unbouded and will lead

to extreme draws. Arguably, it would be preferred to impose this in the estimation procedure, but this would make the estimation procedure substantially more difficult, since it needs to be imposed every time period. Then, the model cannot be easily rewritten into Gaussian state space form and the methods we employ are no longer valid. An alternative is to perform a Metropolis-Hastings step, introducing autocorrelation to the MCMC chain, thereby reducing the sampler's efficiency. Instead, we exclude non-stationary draws when drawing from the predictive density. This implies imposing that $\beta_{4t} < 1$ and applying the same restriction to the autoregressive parameters on the diagonal of $\boldsymbol{\Phi}$. This latter restriction is already imposed in estimation, see Section B. The number of retained draws is 90,460 out of a potential 100,000 for the MI model with unknown break probability π_b.

Conclusions are similar for different condition periods. For brevity, we only present results for conditioning on November 1987. This is one month after Black Monday, a period with relatively high conditional variance, to illustrate high possible differences in time-varying volatility.

The evaluation is in-sample for two reasons. First and foremost, evaluating long-term portfolios out-of-sample is difficult because the sample consists of few non-overlapping periods. This is the reason that most work on long-term allocations is in-sample. For an exception, see e.g. Diris et al. (2014). Second, an out-of-sample analysis is time-consuming. It requires re-estimating the model each period, because we use smoothed estimates. In addition, the predictive density would have to be redrawn each time as well. Johannes et al. (2014) do evaluate out-of-sample. They use filtered estimates and focus on shorter horizons though.

As a measure of the portfolio performance, we use the certainty equivalent return (CER). It is the return that makes the investor indifferent between holding a riskless asset with that return and investing in the (risky) portfolio. In other words, CER is the return the investor is willing to pay to hold the portfolio instead of being fully invested in the risk free asset. So it is the rate that yields the same utility as the portfolio of model i at investment horizon h, computed as $\text{CER}_{ih} = u^{-1}(\text{E}[U_{ih}]) = (\text{E}[U_{ih}](1-\gamma)+1)^{\frac{1}{1-\gamma}} - 1$,

with utility U_{ih}. The CER are annualized assuming continuous compounding.

We are interested in comparing the performance of each model, under various types of parameter stability. Therefore, one model is assumed to be the data generating process (DGP) that generates the predictive density and the CER is computed under that DGP for each of the models with the weights from that particular model. This is similar to the analysis of Ang and Bekaert (2002) and Pettenuzzo and Timmermann (2011), among others. The model of the DGP will always have the highest CER, because the weights are optimized for those draws. Hence, the exercise is repeated assuming each of the models as DGP. As measure for misspecifying the break process, define the loss compared to the optimal CER (i.e. correctly specified time-variation) as $\text{CEL}_{ijh} = \text{CER}_{jjh} - \text{CER}_{ijh}$, the loss for model i at horizon h with model j as the DGP. Performance fees (Fleming et al., 2001) instead of CEL yield qualitatively the same results.

5.2 Term structure of risk

Figure 8 shows the term structure of risk for the MI (with unknown break probability and fixed probability at the expected value of the prior), TVP-VAR, a static VAR model, and MI-b, a MI model where the intercepts and loadings are restricted to be constant. The term structure of risk is the per period standard deviation, so they can be compared across investment horizons. The relatively high short-run variance for the MI and TVP-VAR models reflects the high conditional variance for the starting period. In the medium run, the per period standard deviation decreases due to predictability and mean reversion. This effect is strongest for MI-b and the constant VAR due to stable intercepts and loadings. It indicates that even though MI-b includes time-varying volatility, the effect at the long horizons is mostly driven by the (in)stability of the mean parameters.

In the long run, stocks are riskiest in the MI model. Likely this is because of the additional uncertainty regarding the break probability. The MI model with unknown break probability has a larger variance than the MI model with known break probability. At a 20 years horizon, the variance is 0.25 for the model with fixed $\pi_b = 1/60$ and 0.37 for the model with unknown break probability. This alone does not explain the difference, as

the MI model with known break probability also has higher variance than the TVP-VAR (0.14 at 20 years). There could be stronger mean-reversion in the time-varying parameters for the TVP-VAR model than the MI model. As shown in Figure 3c, there are some posterior draws where the loadings are constant for longer periods. If the parameters are in an extreme regime, and does not revert back to its mean, it can inflate the long run variance.

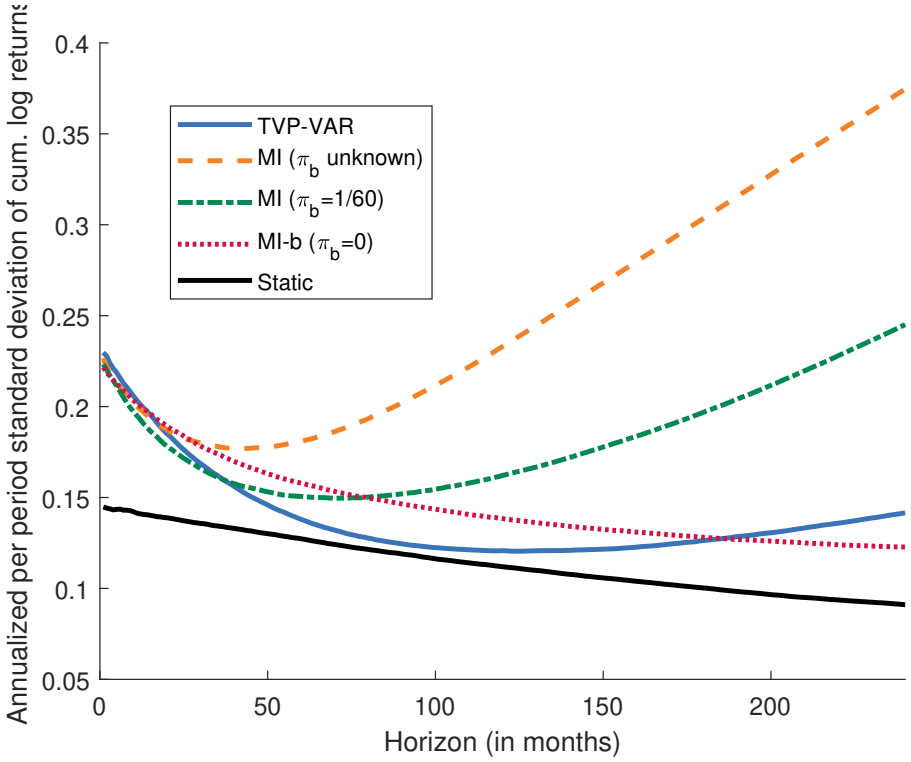
Another possible reason for the difference between the TVP-VAR model and the MI model comes from imposing stationarity and is a bit subtle. The high correlation between the coefficients β_{2t} and β_{4t} implies that by truncating the AR coefficient for the dividend price ratio, β_{4t}, we 'softly truncate' the coefficient of the dividend price ratio on the excess returns, β_{2t}, hence increasing the average predictability coefficient, see Figure 9. If the break probability is high, the predictability dip in the variance is larger, because each period with a break nudges the predictability upwards. This is in favor of the TVP-VAR model and may contribute to the smaller predictive variance for the TVP-VAR model compared to the MI model.

5.3 Weights and certainty equivalent return

Figure 10 presents the fraction of wealth allocated to stocks from optimizing the asset allocation problem in Equation (9) with risk aversion of $\gamma = 5$. Results are qualitatively similar for a risk aversion of 2 and 8. The allocation can be extreme, with the allocations based on the constant VAR and MI-b hitting boundary solutions. When using the constant VAR model to optimize the allocation problem, the allocation is at least 70% and is at the maximum 99% for horizons of 3 years or more. This is a result of the relatively small estimated risk of the returns for those models. When comparing the MI and TVP-VAR, MI has a smaller allocation than TVP-VAR at the horizons longer than 4 years. The weight to the risky asset is 28.3% (MI) and 30.2% (TVP-VAR) at a 20 years horizon. This is due to the large risk at the longer horizon according to the MI model.

The weights show that when allowing for unstable parameters, stocks are less interesting for the long-term investor than when using a static model. However, it does

Figure 8: Term structure of risk

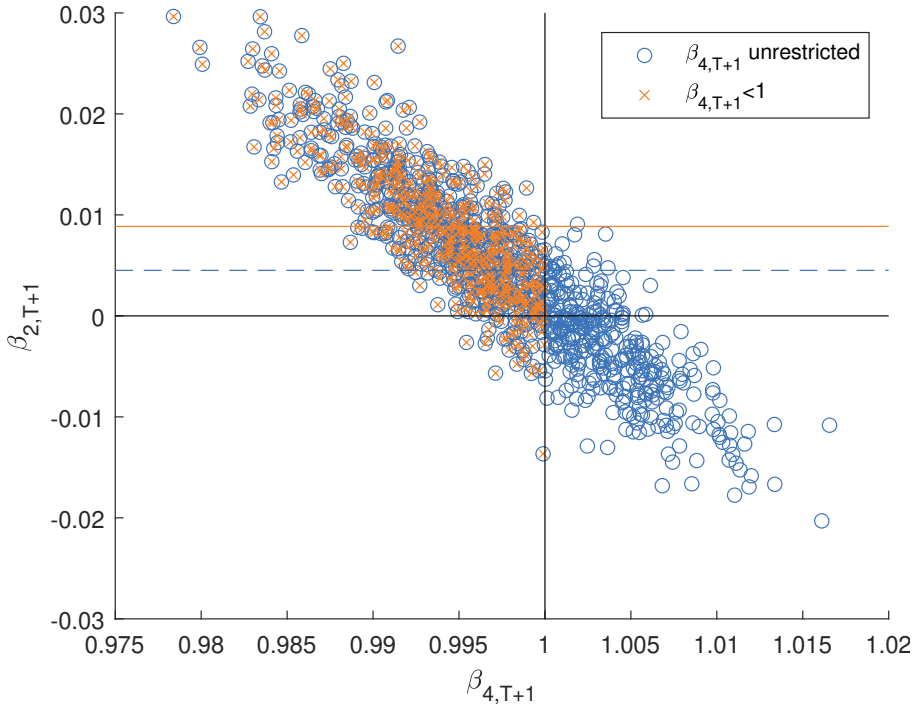


The figure presents the per period standard deviation of the predictive density of stock returns times 12, for an investment horizon up to $h = 240$ months: $\sqrt{12 \times Var_{t+h}(r_t)/h}$. Results are conditional on time period 1987M11.

not tell how much we would lose if we were to assume an incorrect instability of the relationship between the dividend price ratio and excess stock returns. Therefore, we consider the certainty equivalent loss in Table 2. The CERs are annualized and therefore comparable across horizons.

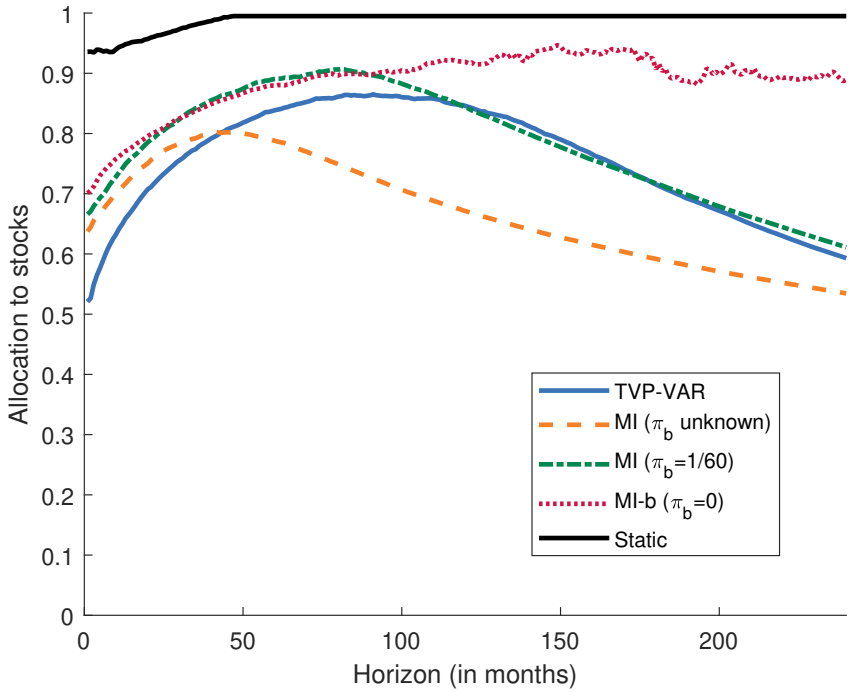
The CELs provide some interesting results. First, assuming constant intercepts and loadings can be costly, especially at the longer horizons. At the short horizon, the losses are limited, but increase as the horizon becomes 5 year or more, up to 16.3% in CER at the 20 year horizon, see panel D in Table 2. Second, vice versa, the MI models and TVP-VAR show a large loss in CER when the DGP has a constant mean. This is because the per period standard deviation is relatively small in the DGP in that case and the investor misses these extra profits. However, it is not too bad for the investor because the absolute CER is still positive. Third, it is important to take parameter stability into account, but the difference between the MI model and the TVP-VAR is limited: 0.02% or 0.04% at the 20 year horizon. Fourth, the impact of time-varying volatility is clear at

Figure 9: Effect of imposing stationarity



The figures present a 1,000 draws from the predictive density for the intercepts and loadings. The blue circles are from the multivariate normal distribution given by the posterior mean and variance for $\boldsymbol{\mu}_1$ (from the static model) and $\boldsymbol{Q}_1$ (from the TVP-VAR model). The orange crosses are in the subset where $|\beta_{4,T+1}| < 1$. The dashed blue line is at the height of the sample mean of $\beta_{2,T+1}$ for the full sample, and the orange solid line is at the height of the sample mean of $\beta_{2,T+1}$ for the subset where $|\beta_{4,T+1}| < 1$.

Figure 10: Allocation to stocks



The figure presents the fraction of wealth allocated to stocks for a buy-and-hold investor with power utility with risk aversion parameter γ is 5. Results are conditional on time period 1987M11.

the short horizon. The CEL in panel A in Table 2 for the static model is much larger than for any other model, even the MI-b. At the long horizon though, instability in the mean dominates, as the allocation and hence CERs for the MI-b and static model are comparable. Fifth, from assuming a DGP with instability uncertainty and considering the CEL for the mixture innovation models without instability uncertainty, the utility cost of ignoring instability uncertainty is only about 0.1% in annualized returns.

Finally, from the bottom row in Table 2, we can identify the model choice that is most robust to the time-variation misspecification, consistent with a min-max utility investor (Hansen and Sargent, 2001). It is the model that maximizes the minimum CER over the various DGPs, where the consideration set is defined by the mixture innovation model we specified, with the break probability prior as choice set. Assuming a TVP-VAR or MI model with $\pi_b = 1/60$ is most beneficial in the long run for the long-term investor with a minimum CER of 4.52% in annualized returns. It's reasonable to allow for any type of instability, with a preference for a high break probability. This might be because the many breaks models allow for more flexibility and don't have to be exact about the break location. Perhaps most important for the long-term investor is the amount of instability and this is well described by these models – also considering their good fit (Section 4.3).

5.4 Comparison to literature

There is a discussion in the literature on whether the predictive variance of stocks increases or decreases with the horizon, see e.g. Pástor and Stambaugh (2012); Carvalho et al. (2018). Traditionally, stocks are believed to be safer in the long run due to mean reversion. However, Pástor and Stambaugh (2012) show that stocks are more volatile in the long run when predictors are assumed to be imperfect proxies of the expected return. Similarly, Pettenuzzo and Timmermann (2011) and Johannes et al. (2014) find that the risk is larger in the long run when taking into account parameter instability. Given that the MI model includes another layer of uncertainty, break risk, it is not surprising that the per period variance increases over the horizon in Figure 8.

In contrast, Carvalho et al. (2018) show in the imperfect predictor setting and with

Table 2: Difference in certainty equivalent return

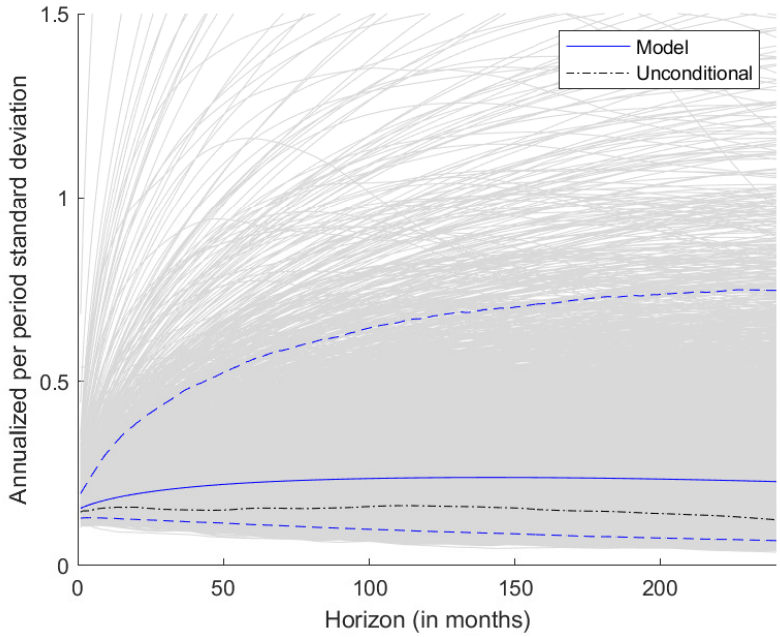
		CEL						
DGP	CER_{DGP}	TVP-VAR	MI π_b unknown	MI $\pi_b = 1/60$	MI $\pi_b = 1/120$	MI $\pi_b = 1/600$	MI-b $\pi_b = 0$	Static
Panel A: $h = 1$								
TVP-VAR	4.85		0.25	0.41	0.46	0.59	0.67	4.94
MI (π_b unknown)	6.07	0.25		0.02	0.03	0.07	0.10	2.96
MI ($\pi_b = 1/60$)	6.42	0.39	0.02		0.00	0.02	0.03	2.44
MI ($\pi_b = 1/120$)	6.57	0.45	0.03	0.00		0.01	0.02	2.36
MI ($\pi_b = 1/600$)	6.82	0.57	0.07	0.02	0.01		0.00	2.06
MI-b ($\pi_b = 0$)	6.97	0.63	0.10	0.03	0.02	0.00		1.93
Static	8.10	2.05	1.24	1.05	1.00	0.88	0.83	
Panel B: $h = 12$								
TVP-VAR	5.72		0.07	0.17	0.18	0.20	0.28	2.40
MI (π_b unknown)	6.48	0.07		0.02	0.02	0.03	0.07	1.61
MI ($\pi_b = 1/60$)	6.81	0.16	0.02		0.00	0.00	0.01	1.21
MI ($\pi_b = 1/120$)	6.90	0.17	0.02	0.00		0.00	0.01	1.21
MI ($\pi_b = 1/600$)	7.19	0.20	0.03	0.00	0.00		0.01	1.26
MI-b ($\pi_b = 0$)	7.51	0.29	0.07	0.01	0.01	0.01		1.05
Static	8.06	1.17	0.79	0.62	0.61	0.58	0.50	
Panel C: $h = 60$								
TVP-VAR	6.53		0.04	0.05	0.04	0.01	0.03	0.82
MI (π_b unknown)	7.01	0.07		0.37	0.33	0.17	0.28	9.89
MI ($\pi_b = 1/60$)	7.48	0.05	0.17		0.00	0.01	0.00	0.30
MI ($\pi_b = 1/120$)	7.59	0.05	0.17	0.00		0.01	0.00	0.53
MI ($\pi_b = 1/600$)	7.93	0.02	0.14	0.03	0.02		0.01	20.97
MI-b ($\pi_b = 0$)	8.40	0.04	0.19	0.00	0.00	0.01		0.75
Static	8.53	0.45	0.70	0.24	0.25	0.34	0.28	
Panel D: $h = 240$								
TVP-VAR	4.52		0.02	0.00	0.04	0.11	1.16	8.79
MI (π_b unknown)	5.40	0.04		0.08	0.29	0.53	3.45	16.28
MI ($\pi_b = 1/60$)	5.43	0.00	0.06		0.05	0.15	2.07	13.69
MI ($\pi_b = 1/120$)	5.72	0.07	0.18	0.04		0.02	1.41	12.64
MI ($\pi_b = 1/600$)	6.55	0.23	0.44	0.17	0.03		1.47	13.63
MI-b ($\pi_b = 0$)	8.69	1.32	1.70	1.21	0.84	0.62		2.21
Static	9.31	2.29	2.69	2.17	1.77	1.51	0.51	
min(CER)		4.52	4.50	4.52	4.47	4.41	1.96	−10.88

The table presents the annualized certainty equivalent loss (CEL), in percentages, the difference in certainty equivalent return between the between assuming a model to choose portfolio weights (columns) and the DGP (rows), and the CER for the DGP, for investment horizon $h \in \{1, 12, 60, 240\}$ in months. Panel D also presents the minimum CER at the 20 year horizon over the DGPs for each of the models. All CERs are for a buy-and-hold investor with power utility and risk aversion of $\gamma = 5$. The conditioning period is 1987M11. Results are based on up to 100,000 draws, excluding non-stationary draws.

"plausible priors" that the predictive variance can be smaller in the long run. They argue that Johannes et al. (2014) overestimate the risk. To assess if this is true for the MI model, we repeat their exercise by comparing draws from the predictive density to the "unconditional average," the sample average based on h period returns. We deviate

from Carvalho et al. (2018) and compute the variances using Newey-West with $h-1$ lags to account for overlapping periods. Figure 11 shows that the model average does not track the sample average exactly, but it is well in the credible interval. So the model describes the data reasonably well. This does not contradict Carvalho et al. (2018). In their appendix, they also find that for monthly returns the predictive variance increases over the horizon. For lower frequency data they do conclude that stocks are safer in the long run.

Figure 11: Model implied and unconditional predictive variance



The figure presents draws (gray), mean (solid blue) and 90% credible interval (dashed blue) of the per period standard deviation of the predictive density of stock returns times 12, for an investment horizon up to $h = 240$ months: $\sqrt{12 \times Var_{t+h}(r_t)/h}$, for the MI model with unknown break probability. Results are conditional on time period 1987M11. The dash-dotted black line is the sample average, computed from all h period returns, using Newey-West with $h-1$ bandwidth.

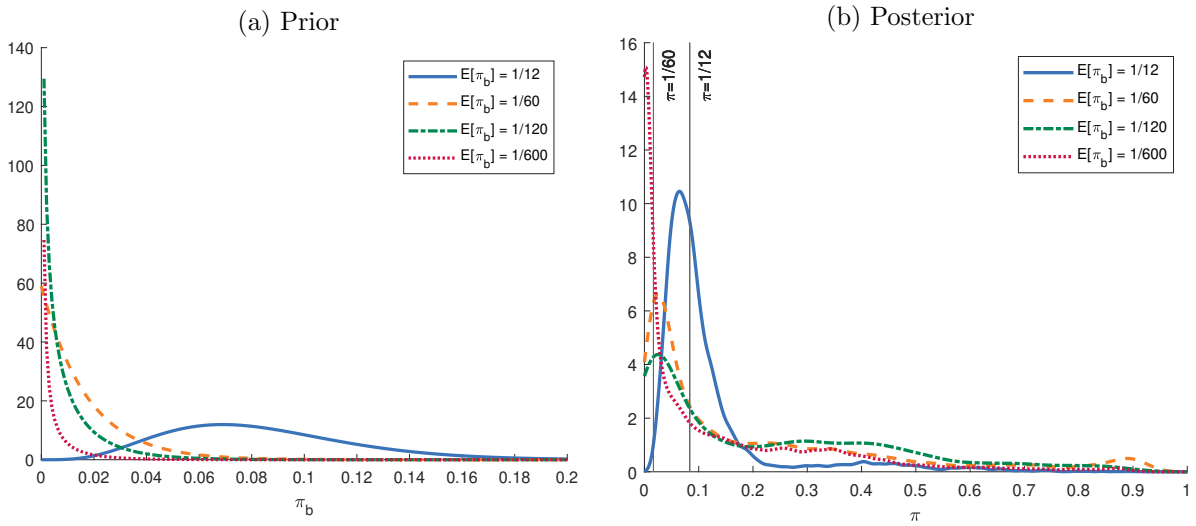
6 Prior sensitivity analysis

To assess the sensitivity of the results to our prior choice, we re-estimate the MI model, with different priors for the break probability π_b, the parameter of interest. The hyperparameters are chosen such that the prior strength is equal to that of the baseline prior ($\underline{a}_b = 1$ and $\underline{b}_b = 59$), and matches the mean to each of the considered break probabilities in Section 4.1. For example, to get an expected value of $\pi_b = 1/120$, $\underline{a}_b = 0.5$

and $\underline{b}_b = 59.5$. The different priors are plotted in Figure 12a.

The posterior of π_b in Figure 12b shows some of the prior influence in that the modes are ordered as expected. The mode is 0.3% for the very few breaks prior ($\mathrm{E}[\pi_b = 1/600]$) to 6.5% for the many breaks prior ($\mathrm{E}[\pi_b = 1/12]$). The posterior is wide for all different priors, and the probability of many breaks ($\pi > 1/12$) is substantial, ranging from 40% ($\mathrm{E}[\pi_b = 1/600]$) to 57% ($\mathrm{E}[\pi_b = 1/120]$). This indicates considerable uncertainty regarding the break probability. Figure 13 shows that the general pattern of the time-varying parameters is quite robust to the prior values. The variance is the same for all models, and the time-variation in β_{2t} is of a similar magnitude.

Figure 12: Distribution of π_b



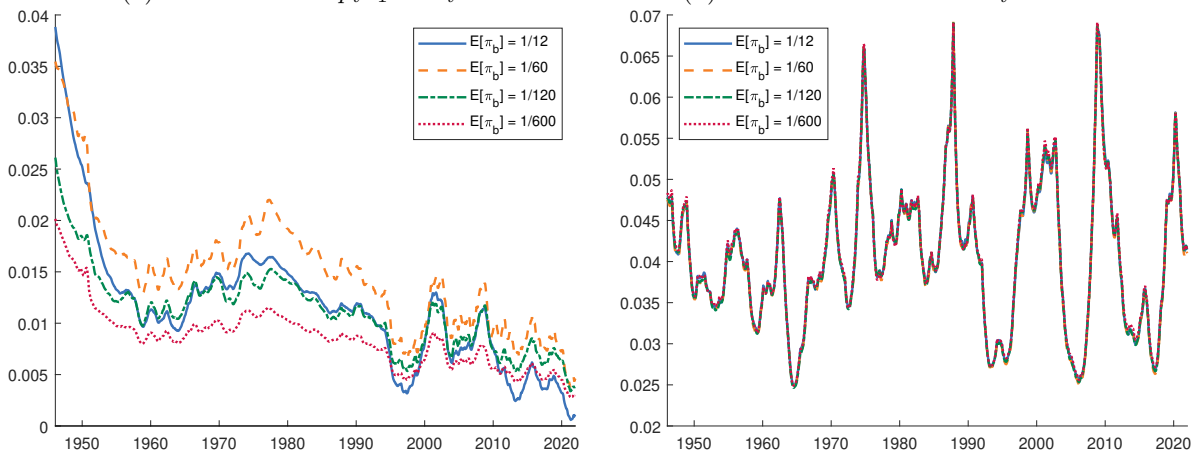
The figures present the prior and posterior density for the break probability of the intercepts and loadings for mixture innovation models with prior expected values for the break probability equal to $\mathrm{E}[\pi_b] \in \{1/600, 1/120, 1/60, 1/12\}$, with hyperparameters $\underline{a}_b + \underline{b}_b = 60$. The hyperparameters for the break probability of the variances and the covariance term are set at $\underline{a}_s = \underline{b}_s = \underline{a}_a = \underline{b}_a = 1$.

More relevant is to assess the impact of the prior on the predictive variance and allocation. The predictive variance in Figure 14a shows that it can vary a bit across the different priors, ranging at the 20 years horizon from 0.21 ($\mathrm{E}[\pi_b] = 1/12$) to 0.39 ($\mathrm{E}[\pi_b] = 1/120$). However, Figure 14 shows that the allocation is very similar for all prior specifications, which is also clear from the small losses in Table 3. The CEL in Table 3 ranges from 0 up to 36 basis points, so the costs of misspecifying the prior are limited. In that sense, the results are robust to the prior specification.

The effect of instability uncertainty is consistent across different priors. The added

Figure 13: Posterior mean of time-varying parameters for different priors of π_b

(a) Coefficient of dp_{t-1} on r_t (b) Standard deviation of r_t residuals



The figures present the posterior mean the predictability coefficient, β_{2t}, and the residual standard deviation of r, σ_{1t}, from the mixture innovation model in Equations (1)–(2) with the time-varying parameters described by Equations (4)–(7). If π_b is assumed unknown, the priors are set such that the expected value equal is to $\mathrm{E}[\pi_b] \in \{1/600, 1/120, 1/60, 1/12\}$, with hyperparameters $\underline{a}_b + \underline{b}_b = 60$. The hyperparameters for the break probability of the variances and the covariance term set at $\underline{a}_s = \underline{b}_s = \underline{a}_a = \underline{b}_3 = a$. If the break probability is assumed known, it is fixed at the prior mean, such that $\pi_b \in \{1/600, 1/120, 1/60, 1/12\}$ and $\pi_s = \pi_a = 1$. See Section 2.3 for the full prior specification.

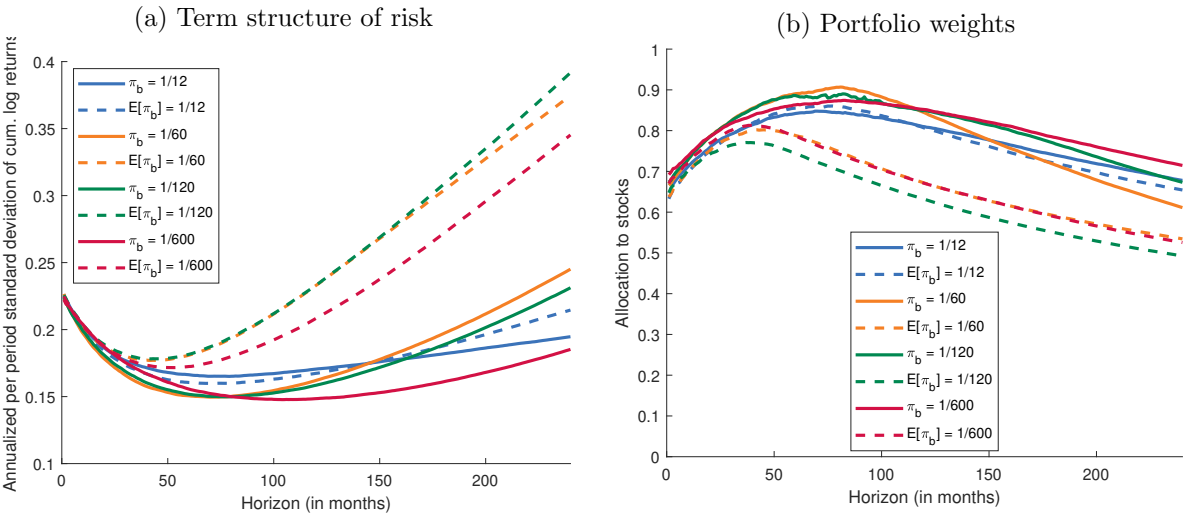
uncertainty widens the predictive density (Figure 14a) and this extra risk translates to a lower allocation to stocks (Figure 14b). The uncertainty accumulates over the horizon as the differences are larger as the horizon increases. Utility costs of ignoring break probability uncertainty for an investor with a 20 year horizon vary from 0.01% for a prior at many breaks to 0.66% for lower break probabilities, see the last column in Table 3. So these are limited, but can be non-negligible.

Table 3: CEL for different priors of π_b

	$\mathrm{E}[\pi_b] = 1/12$	$\mathrm{E}[\pi_b] = 1/60$	$\mathrm{E}[\pi_b] = 1/120$	$\mathrm{E}[\pi_b] = 1/600$	$\pi_b = \mathrm{E}[\pi_b]$
$\mathrm{E}[\pi_b] = 1/12$		0.15	0.25	0.17	0.01
$\mathrm{E}[\pi_b] = 1/60$	0.21		0.02	0.00	0.08
$\mathrm{E}[\pi_b] = 1/120$	0.36	0.02		0.01	0.47
$\mathrm{E}[\pi_b] = 1/600$	0.28	0.00	0.01		0.66

The table presents the annualized certainty equivalent loss (CEL), in percentages, the difference in certainty equivalent return between assuming a mixture innovation model to choose portfolio weights (columns) and the DGP (rows), for an investment horizon of $h = 240$ months. The final column presents the difference in CER from assuming a mixture innovation model with known break probability equal to the prior expected value as DGP. All CERs are for a buy-and-hold investor with power utility and risk aversion of $\gamma = 5$. The conditioning period is 1987M11.

Figure 14: Asset allocation for different priors of π_b



The figures present (a) the per period standard deviation of the predictive density of stock returns times 12, $\sqrt{12 \times Var_{t+h}(r_t)/h}$, and (b) the allocation to stocks for a buy-and-hold investor with power utility with risk aversion parameter γ is 5, for an investment horizon up to $h = 240$ months. Results are conditional on time period 1987M11.

7 Conclusion

We assess the economic costs of misspecifying the type of instability, few (large) or many (small) breaks, in the predictive relationship between the dividend price ratio and stock returns for a long-term investor. We estimate a mixture innovation model with Bayesian estimation methodology to quantify the effect of uncertainty regarding the break probability. The predictive relationship seems to be subject to many breaks rather than few breaks. There is substantial uncertainty though, likely due to the low signal-to-noise ratio. The uncertainty adds to the volatility in the predictive density and hence lower allocation to stocks.

From a long-term investor's perspective, the costs of ignoring parameter instability are high, even if the true process is subject to a limited number of breaks. The costs can run up to 16% in terms of annualized certainty equivalent return over a 20 year horizon. Conditional on assuming instabilities, the costs of misspecifying the instability are limited. Therefore, long-term investors should at least allow for the possibility of parameter instability. Allowing for more breaks is preferred, which probably helps because of the uncertainty on the number and location of breaks.

References

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

Ang, A. and Bekaert, G. (2007). Stock return predictability: Is it there? *Review of Financial Studies*, 20(3):651–707.

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

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

Bauwens, L., Koop, G., Korobilis, D., and Rombouts, J. V. (2015). The contribution of structural break models to forecasting macroeconomic series. *Journal of Applied Econometrics*, 30:596—-620.

Brandt, M. (2010). Portfolio choice problems. In Aït-Sahalia, Y. and Hansen, L. P., editors, *Handbook of financial econometrics*, volume 1, chapter 5, pages 269–336. North Holland.

Campbell, J. Y. and Shiller, R. J. (1988). The dividend-price ratio and expectations of future dividends and discount factors. *Review of Financial Studies*, 1(3):195–228.

Campbell, J. Y. and Viceira, L. M. (2002). *Strategic asset allocation: Portfolio choice for long-term investors*. Oxford University Press.

Carter, C. K. and Kohn, R. (1997). Semiparametric Bayesian inference for time series with Mixed spectra. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 59(1):255–268.

Carvalho, C. M., Lopes, H. F., and McCulloch, R. E. (2018). On the long-run volatility of stocks. *Journal of the American Statistical Association*, 113(523):1050–1069.

Chib, S. (1998). Estimation and comparison of multiple change-point models. *Journal of Econometrics*, 86(2):221–241.

Clark, T. E. and Ravazzolo, F. (2015). Macroeconomic forecasting performance under alternative specifications of time-varying volatility. *Journal of Applied Econometrics*, 30(4):551–575.

Cogley, T. and Sargent, T. J. (2001). Evolving post-World War II US inflation dynamics. *NBER Macroeconomics Annual*, 16:331–373.

Cogley, T. and Sargent, T. J. (2005). Drifts and volatilities: monetary policies and outcomes in the post WWII US. *Review of Economic Dynamics*, 8(2):262–302.

Cremers, K. J. M. (2002). Stock return predictability: A Bayesian model selection perspective. *Review of Financial Studies*, 15(4):1223–1249.

Dangl, T. and Halling, M. (2012). Predictive regressions with time-varying coefficients. *Journal of Financial Economics*, 106(1):157–181.

Del Negro, M. and Primiceri, G. E. (2015). Time varying structural vector autoregressions and monetary policy: A corrigendum. *Review of Economic Studies*, 82(4):1342–1345.

Diris, B. (2014). The impact of model instability on long-term investors. Available at SSRN: http://ssrn.com/abstract=2480046.

Diris, B., Palm, F., and Schotman, P. (2014). Long-term strategic asset allocation: An out-of-sample evaluation. *Management Science*, 61(9):2185–2202.

Durbin, J. and Koopman, S. J. (2002). A simple and efficient simulation smoother for state space time series analysis. *Biometrika*, 89(3):603–616.

Durbin, J. and Koopman, S. J. (2012). *Time series analysis by state space methods*. Oxford University Press.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4):987–1007.

Farmer, L. E., Schmidt, L., and Timmermann, A. (2023). Pockets of predictability. *The Journal of Finance*, 78(3):1279–1341.

Fiorentini, G., Planas, C., and Rossi, A. (2014). Efficient MCMC sampling in dynamic mixture models. *Statistics and Computing*, 24(1):77–89.

Fleming, J., Kirby, C., and Ostdiek, B. (2001). The economic value of volatility timing. *The Journal of Finance*, 56(1):329–352.

Gelman, A., Hwang, J., and Vehtari, A. (2014). Understanding predictive information criteria for Bayesian models. *Statistics and Computing*, 24(6):997–1016.

Gerlach, R., Carter, C., and Kohn, R. (2000). Efficient Bayesian inference for dynamic mixture models. *Journal of the American Statistical Association*, 95(451):819–828.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In Bernardo, J. M., Berger, J. O., Dawid, A. P., and Smith, A. F. M., editors, *Bayesian statistics*, volume 4, pages 169–194. Oxford University Press, Oxford.

Giordani, P. and Kohn, R. (2008). Efficient Bayesian inference for multiple change-point and mixture innovation models. *Journal of Business & Economic Statistics*, 26(1):66–77.

Groen, J. J., Paap, R., and Ravazzolo, F. (2013). Real-time inflation forecasting in a changing world. *Journal of Business & Economic Statistics*, 31(1):29–44.

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

Hansen, L. P. and Sargent, T. J. (2001). Robust control and model uncertainty. *American Economic Review*, 91(2):60–66.

Henkel, S. J., Martin, J. S., and Nardari, F. (2011). Time-varying short-horizon predictability. *Journal of Financial Economics*, 99(3):560–580.

Johannes, M., Korteweg, A., and Polson, N. (2014). Sequential learning, predictability, and optimal portfolio returns. *Journal of Finance*, 69(2):611–644.

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

Kim, S., Shephard, N., and Chib, S. (1998). Stochastic volatility: Likelihood inference and comparison with ARCH Models. *Review of Economic Studies*, 65(3):361–393.

Koop, G., Leon-Gonzalez, R., and Strachan, R. W. (2009). On the evolution of the monetary policy transmission mechanism. *Journal of Economic Dynamics and Control*, 33(4):997–1017.

Koop, G. and Potter, S. M. (2007). Estimation and forecasting in models with multiple breaks. *Review of Economic Studies*, 74(3):763–789.

Lettau, M. and Van Nieuwerburgh, S. (2008). Reconciling the return predictability evidence. *Review of Financial Studies*, 21(4):1607–1652.

McCulloch, R. E. and Tsay, R. S. (1993). Bayesian inference and prediction for mean and variance shifts in autoregressive time series. *Journal of the American Statistical Association*, 88(423):968–978.

Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708.

Pástor, L. and Stambaugh, R. F. (2012). Are stocks really less volatile in the long run? *Journal of Finance*, 67(2):431–478.

Paye, B. S. and Timmermann, A. (2006). Instability of return prediction models. *Journal of Empirical Finance*, 13(3):274–315.

Pesaran, M. H. and Timmermann, A. (1995). Predictability of stock returns: Robustness and economic significance. *Journal of Finance*, 50(4):1201–1228.

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

Pettenuzzo, D. and Timmermann, A. (2017). Forecasting macroeconomic variables under model instability. *Journal of Business & Economic Statistics*, 35(2):183–201.

Primiceri, G. E. (2005). Time varying structural vector autoregressions and monetary policy. *Review of Economic Studies*, 72(3):821–852.

Rapach, D. and Zhou, G. (2013). Forecasting stock returns. In Elliott, G. and Timmermann, A., editors, *Handbook of economic forecasting*, volume 2A, chapter 6, pages 328–383. Elsevier Amsterdam.

Rapach, D. E. and Wohar, M. E. (2006). Structural breaks and predictive regression models of aggregate US stock returns. *Journal of Financial Econometrics*, 4(2):238–274.

Ravazzolo, F., Paap, R., van Dijk, D., and Franses, P. H. (2008). Bayesian model averaging in the presence of structural breaks. In Rapach, D. E. and Wohar, M. E., editors, *Forecasting in the presence of structural breaks and model uncertainty*, chapter 15, pages 561–594. Emerald Group Publishing Limited.

Stroud, J. R., Müller, P., and Polson, N. G. (2011). Nonlinear state-space models with state-dependent variances. *Journal of the American Statistical Association*, 98(462):377–386.

Timmermann, A. (2008). Elusive return predictability. *International Journal of Forecasting*, 24(1):1–18.

Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, 11(Dec):3571–3594.

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

Appendix A ARMI process properties

Define the autoregressive mixture innovation (ARMI) process X_t as

$$\begin{aligned} X_t &= \mu + f(\kappa_t, \phi)(X_{t-1} - \mu) + \kappa_t \eta_t \\ &= \mu + \big(\phi\kappa_t + (1 - \kappa_t)\big)(X_{t-1} - \mu) + \kappa_t \eta_t, \end{aligned}$$

with $\eta_t \sim \mathrm{N}(0, \sigma^2)$ and $\Pr[\kappa_t = 1] = \pi$.

A.1 Stationarity conditions

First, we show that the process X_t is stationary if $|\phi| < 1$. Without loss of generality, we take $\mu = 0$ for notation convenience. We rewrite X_t as a sum of the initial condition and its innovations,

$$\begin{aligned} X_t &= X_0 \prod_{s=1}^{t} (\phi\kappa_s + (1 - \kappa_s)) + \sum_{s=1}^{t} \left(\prod_{j=s+1}^{t} (\phi\kappa_j + (1 - \kappa_j)) \right) \kappa_s \eta_s \\ &= X_0 \prod_{s=1}^{t} f(\kappa_s, \phi) + \sum_{s=1}^{t} \left(\prod_{j=s+1}^{t} f(\kappa_j, \phi) \right) \kappa_s \eta_s. \end{aligned}$$

This sum is non-explosive if $|f(\kappa_t, \phi)| < 1$ and $|[\prod_{j=s+1}^{t} f(\kappa_j, \phi)]\kappa_s| < 1$. This depends on the break process $\boldsymbol{\kappa}$. If we assume, without loss of generality, that a fraction π^* periods for which $\kappa_s = 1$ for $1 \leq s \leq t$ and a fraction $\tilde{\pi}$ periods for which $\kappa_j = 1$ for $s+1 \leq j \leq t$, then

$$\begin{aligned} \prod_{s=1}^{t} f(\kappa_s, \phi) &= \prod_{s=1}^{t} (\phi\kappa_s + (1 - \kappa_s)) \\ &= \left(\prod_{1 \leq s \leq t | \kappa_s = 1} (\phi\kappa_s + (1 - \kappa_s)) \right) \left(\prod_{1 \leq s \leq t | \kappa_s = 0} (\phi\kappa_s + (1 - \kappa_s)) \right) \\ &= \left(\prod_{1 \leq s \leq t | \kappa_s = 1} \phi \right) \left(\prod_{1 \leq s \leq t | \kappa_s = 0} 1 \right) \\ &= \phi^{\pi^* t}, \end{aligned}$$

and

$$\left(\prod_{j=s+1}^{t} f(\kappa_j, \phi)\right)\kappa_s = \left(\prod_{j=s+1}^{t} (\phi\kappa_j + (1-\kappa_j))\right)\kappa_s = \phi^{\tilde{\pi}(t-s)}\kappa_s.$$

Since $0 \leq \pi^* \leq 1$, it follows that $|\phi^{\pi^* t}| < 1$ and $|\phi^{\tilde{\pi}(t-s)}\kappa_s| \leq 1$ if $|\phi| < 1$. Strict inequality holds for both, unless $\tilde{\pi} = 0$. Then the first part becomes one and we are left with κ_s, which is either zero or one. However, $\kappa_s\eta_s$ is not explosive if $\kappa_s = 1$, as it is only a single normally distributed random variable, not a sum. Moreover, the realization of η_s will then be discounted for future values X_j where $j > s$ if $|\phi| < 1$. This shows that the ARMI process X_t is stationary if $|\phi| < 1$.

A.2 Unconditional distribution

We show that the unconditional distribution of the ARMI process given by Equation (4), under stationarity, i.e. $|\phi| < 1$, and $\Pr[\kappa_t = 1] = \pi > 0$, is the same as for a stationary autoregressive process, i.e. if $\pi = 1$. This is equivalent to proving that the unconditional distribution does not depend on the break probability π. We assume $|\phi| < 1$, $\pi > 0$ and independence between X_{t-1} and κ_t.

The unconditional mean of X_t is

$$\begin{aligned}
\mathrm{E}[X_t] &= \mathrm{E}\left[\mu + \kappa_t\phi(X_{t-1} - \mu) + (1-\kappa_t)(X_{t-1} - \mu) + \kappa_t\eta_t\right] \\
&= \mu + \mathrm{E}\left[\kappa_t\phi(X_{t-1} - \mu)\right] + \mathrm{E}\left[(1-\kappa_t)(X_{t-1} - \mu)\right] \\
\mathrm{E}[X_t - \mu] &= \mathrm{E}[\kappa_t]\phi\,\mathrm{E}[X_{t-1} - \mu] + \mathrm{E}[1-\kappa_t]\,\mathrm{E}[X_{t-1} - \mu] \\
&= \pi\phi\,\mathrm{E}[X_t - \mu] + (1-\pi)\,\mathrm{E}[X_t - \mu] \\
\big(1 - \pi\phi - (1-\pi)\big)\,\mathrm{E}[X_t - \mu] &= 0 \\
\mathrm{E}[X_t - \mu] &= 0 \\
\mathrm{E}[X_t] &= \mu.
\end{aligned}$$

The unconditional variance of X_t is

$$\begin{aligned}
\operatorname{var}(X_t - \mu) &= \operatorname{var}\big(\mu + \kappa_t\phi(X_{t-1} - \mu) + (1 - \kappa_t)(X_{t-1} - \mu) + \kappa_t\eta_t\big) \\
&= \phi^2 \operatorname{var}\big(\kappa_t(X_{t-1} - \mu)\big) + \operatorname{var}\big((1 - \kappa_t)(X_{t-1} - \mu)\big) + \operatorname{var}(\kappa_t\eta_t) \\
&= \phi^2 \operatorname{E}[\kappa_t] \operatorname{var}(X_{t-1} - \mu) + \operatorname{E}[1 - \kappa_t] \operatorname{var}(X_{t-1} - \mu) + \operatorname{E}[\kappa_t] \operatorname{var}(\eta_t) \\
&= \phi^2\pi \operatorname{var}(X_t - \mu) + (1 - \pi) \operatorname{var}(X_t - \mu) + \pi\sigma^2 \\
\big(1 - \phi^2\pi - (1 - \pi)\big) \operatorname{var}(X_t - \mu) &= \pi\sigma^2 \\
(\pi - \phi^2\pi) \operatorname{var}(X_t - \mu) &= \pi\sigma^2 \\
\operatorname{var}(X_t - \mu) &= \frac{\pi\sigma^2}{\pi - \phi^2\pi} = \frac{\sigma^2}{1 - \phi^2}.
\end{aligned}$$

Hence, the unconditional distribution of X_t is $\mathrm{N}\big(\mu, \sigma^2/(1 - \phi^2)\big)$ and does not depend on the break probability π.

Appendix B Priors

We employ Bayesian methodology to estimate the MI model in Equations (1)–(2) and Equations (4) and (7). This is a natural way to include parameter uncertainty, a relevant risk for the long-term investor (Barberis, 2000). Moreover, it allows us to assess the degree of uncertainty on the break probability. Given the Bayesian methodology, we need to specify prior distributions for the static parameters and initial conditions for the state equations. The priors are mostly standard conjugate empirically Bayesian priors, based on Primiceri (2005) and Koop et al. (2009). The location and scaling are based on ordinary least squares (OLS) estimates of the restricted VAR in Equations (1)–(2) using the first ten years of data. This period is excluded from the estimation sample. We use fairly informative priors regarding the initial values of the state variables, but less so for the long run mean and break probability.

As prior for the long run means $\boldsymbol{\mu}$, we take a normal distribution,

$$\boldsymbol{\mu}_b \sim \mathrm{N}(\hat{\boldsymbol{\beta}}, 10^6 \operatorname{var}(\hat{\boldsymbol{\beta}})), \tag{B.1}$$

$$\boldsymbol{\mu}_s \sim \mathrm{N}(\widehat{\log \boldsymbol{\sigma}^2}, 10^6 \operatorname{var}(\widehat{\log \boldsymbol{\sigma}^2})), \tag{B.2}$$

$$\boldsymbol{\mu}_a \sim \mathrm{N}(\hat{\alpha}, 10^6 \operatorname{var}(\hat{\alpha})), \tag{B.3}$$

where $\hat{\boldsymbol{\beta}}$ is the OLS estimator on the first ten years of data and $\operatorname{var}(\hat{\boldsymbol{\beta}})$ its variance covariance matrix. To ensure that the prior is uninformative, the variance is multiplied by a large number, 10^6. We do not have OLS estimates for the structural log variances $\log \boldsymbol{\sigma}^2$ and covariance term α and their variances, so we follow Koop et al. (2009) and draw a 1,000 times from an inverse Wishart distribution,

$$\boldsymbol{\Omega}^* \sim \mathrm{IW}(\hat{\boldsymbol{\Omega}}\tau, \tau), \tag{B.4}$$

with $\hat{\boldsymbol{\Omega}}$ the OLS estimate of the residual covariance matrix and degrees of freedom $\tau = 120$ set to the number of time periods used to calibrate the prior. Note that the degrees of freedom in an inverse Wishart distribution can indeed be interpreted as the number of

observations to calibrate the location. For each draw of $\boldsymbol{\Omega}^*$, we use the decomposition $\boldsymbol{A}_t\boldsymbol{\Omega}_t\boldsymbol{A}_t' = \boldsymbol{\Sigma}_t\boldsymbol{\Sigma}_t'$ to obtain a draw of the structural log variances $\log\boldsymbol{\sigma}^2$ and covariance term α. The sample mean and variance of the draws are the estimates $\widehat{\log\boldsymbol{\sigma}^2}$ and $\hat{\alpha}$ and their variances $\text{var}(\widehat{\log\boldsymbol{\sigma}^2})$ and $\text{var}(\hat{\alpha})$.

The off-diagonal elements of $\boldsymbol{\Phi}_k$ are set to zero and we take a truncated normal distribution as prior for the diagonal of the autoregressive parameter matrix $\boldsymbol{\Phi}_k$,

$$\text{diag}(\boldsymbol{\Phi}_k) \sim \text{N}(\underline{\boldsymbol{m}}_k, \underline{\boldsymbol{V}}_k)I(|\phi_{ki}| < 1, \forall i = 1, \ldots, n_k), \tag{B.5}$$

for $k \in \{b, s, a\}$, where $I(A)$ is one if condition A holds and zero otherwise, with $\underline{m}_{ki} = 0.9$ and $\underline{V}_{k(i,i)} = 0.2^2$ and $\underline{V}_{k(i,j)} = 0$ for $i, j = 1, \ldots, n_k$, $i \neq j$ and $k \in \{b, s, a\}$. The prior is truncated to ensure stationary parameters, as discussed in Section 2.2. We use the notation of adding an underscore bar to prior hyperparameters. The mean and variance reflect our belief in a persistent process for the parameters.

As prior for the break probability π_k, we take a beta distribution,

$$\pi_k \sim \text{Beta}(\underline{a}_k, \underline{b}_k), \tag{B.6}$$

for $k \in \{b, s, a\}$. The hyperparameters $\underline{a}_k$ and $\underline{b}_k$ reflect the investor's views on the presence of breaks (or lack thereof). For the intercepts and loadings, we set the hyperparameters $\underline{a}_k$ and $\underline{b}_k$ such that the prior's expectation is between a few breaks (about one break every ten years) and many breaks (about one break per year) case. In particular, we set $\underline{a}_b = 1$ and $\underline{b}_b = 59$, which implies an average duration of 60 months between breaks, and implies a fairly weak prior. The hyperparameters in the beta prior can be thought of as adding $\underline{a} + \underline{b}$ observations: $\underline{a}$ ones and $\underline{b}$ zeros, or $\underline{a}$ breaks. If we define the prior strength as $(\underline{a}+\underline{b})/T$, with T the sample size, we have a prior strength of 7.1%. Ravazzolo et al. (2008) and Groen et al. (2013) employ priors with prior strengths of 7.4% and 5.2%.

For the variances and covariance's break processes, we take $\underline{a}_s = \underline{b}_s = \underline{a}_a = \underline{b}_a = 1$. It is uninformative in π, but at the same time puts most probability mass at the

many breaks case, to favor stochastic volatility. Ignoring the stylized fact of volatility clustering might bias the amount of time-variation in the loadings upwards. We have also experimented with setting the hyperparameters to an intermediate break process for all break processes, i.e. $\underline{a}_k = 1$ and $\underline{b}_k = 59$, for $k \in \{b, s, a\}$. This would be in line with for example Pettenuzzo and Timmermann (2011), who estimate a change point model with simultaneous breaks in the mean and the variance. The posterior estimates are implausible in that they imply a (nearly) constant variance, but volatile loadings. Time-variation is then mostly explained by the mean, because this has the largest impact on the fit. This is likely due to the low signal-to-noise ratio in the data.

As prior for the break size matrix $\boldsymbol{Q}_k$ we have an inverse Wishart distribution,

$$\boldsymbol{Q}_k \sim \text{IW}(\underline{\boldsymbol{W}}_k \underline{\nu}_k, \underline{\nu}_k), \tag{B.7}$$

for $k \in \{b, s, a\}$, with $\underline{\nu}_k$ degrees of freedom and

$$\underline{\boldsymbol{W}}_b = \frac{c_b}{\text{E}[\pi_b]} \text{var}(\hat{\boldsymbol{\beta}}), \tag{B.8}$$

$$\underline{\boldsymbol{W}}_s = \frac{c_s}{\text{E}[\pi_s]} \text{var}(\widehat{\log \boldsymbol{\sigma}^2}), \tag{B.9}$$

$$\underline{w}_a = \frac{c_a}{\text{E}[\pi_a]} \text{var}(\hat{\alpha}), \tag{B.10}$$

where c_k is a scaling factor, $\text{E}[\pi_k]$ is the expected prior break probability, and $\text{var}(\hat{\boldsymbol{\beta}})$, $\text{var}(\widehat{\log \boldsymbol{\sigma}^2})$, and $\text{var}(\hat{\alpha})$ are the variance covariance matrices of the OLS estimators of $\boldsymbol{\beta}$, $\log \boldsymbol{\sigma}^2$, and α on the first ten years of data as in Equations (B.1)–(B.3).

The model is known to be sensitive to the choice of c_k, see e.g. Primiceri (2005) for a discussion. We set $c_k = 0.001$ for $k \in \{b, s, a\}$, but setting $c_k = 0.01$ or $c_k = 0.0001$ does not affect the results qualitatively.

The presence of $\text{E}[\pi_k]$ in Equations (B.8)–(B.10) reflects the assumption that the time-variation can be characterized either as a small number of large breaks or a large number of small breaks. Following Koop et al. (2009), if the prior for the break probability changes, the prior for the break size should change accordingly.

We should be careful when choosing the degrees of freedom $\underline{\nu}_k$. To see why, consider the case where π_1 is close to 0. The full conditional posterior for the break size $\boldsymbol{Q}_b$ is an inverse Wishart distribution,

$$\boldsymbol{Q}_b|\ldots \sim \text{IW}\left(\underline{\boldsymbol{W}}_b\underline{\nu}_b + \sum_{t=2}^{T}\kappa_{bt}\boldsymbol{\eta}_t\boldsymbol{\eta}_t', \underline{\nu}_b + \sum_{t=2}^{T}\kappa_{bt}\right), \tag{B.11}$$

with κ_{bt} the break process and $\boldsymbol{\eta}_t$ the residuals from Equation (4) of the current draw, and the notation "$|\ldots$" denotes conditioning on the data and all other parameters. Because π_b is close to zero, it is likely that the number of breaks in the current draw is small or even zero. Therefore, $\sum_{t=2}^{T}\kappa_{1t} \approx 0$ and the break size posterior is (approximately) equal to the prior. This has two consequences. First, the prior degrees of freedom needs to be sufficiently large to ensure that the (finite) break size variance exists, i.e. $\underline{\nu}_k > n_k+3$, with n_k the matrix diagonal's size. Actually, we need more degrees of freedom, because even if the variance exists, for small degrees of freedom, the posterior break size variance still explodes if the break probability is small and inference is complicated due to unrealistic draws. Second, if the break probability π_k is small, the break size prior is very informative, even if the prior degrees of freedom $\underline{\nu}_k$ is small. This is because if π_b is small, the sum $\sum_{t=2}^{T}\kappa_{bt}\boldsymbol{\eta}_t\boldsymbol{\eta}_t'$ will be close to zero, and the mean of $\boldsymbol{Q}_b$ is (almost) fully determined by the prior's location $\underline{\boldsymbol{W}}_b\underline{\nu}_b$. Hence, one should be careful when interpreting the break size results in that case. In a TVP model, the degrees of freedom is often set equal to the minimum for the (finite) prior break size variance to exist. Instead of one, we add ten degrees of freedom, such that $\underline{\nu}_k = n_k + 3 + 10$, to limit the prior's informativeness, while ensuring that the break size variance does not explode if the break probability is small.

As priors for the initial conditions of the state equations we take a normal distribution,

$$\boldsymbol{\beta}_1 \sim \text{N}(\boldsymbol{0}, \underline{\nu}_b \,\text{var}(\hat{\boldsymbol{\beta}})), \tag{B.12}$$

$$\log \boldsymbol{\sigma}_1^2 \sim \text{N}(\boldsymbol{0}, \underline{\nu}_s \,\text{var}(\widehat{\log \boldsymbol{\sigma}^2})), \tag{B.13}$$

$$\alpha_1 \sim \text{N}(0, \underline{\nu}_a \,\text{var}(\hat{\alpha})), \tag{B.14}$$

where $\text{var}(\hat{\boldsymbol{\beta}})$, $\text{var}(\widehat{\log \boldsymbol{\sigma}^2})$, and $\text{var}(\hat{\alpha})$ are the variance covariance matrices of the OLS

estimators of $\boldsymbol{\beta}$, $\log \boldsymbol{\sigma}^2$, and α on the first ten years of data as in Equations (B.1)–(B.3), and $\underline{\nu}_b$, $\underline{\nu}_s$, and $\underline{\nu}_a$ equal to the degrees of freedom in the prior of the break size in Equations (B.8)–(B.10). This is quite informative, but consistent with that the initial value is likely close to the OLS estimate of the ten preceding years.

Appendix C Gibbs sampler

The Gibbs sampler is closely related to the one used by Koop et al. (2009). The difference is the addition of autoregressive parameters and long run means. The sampler is based on the algorithm for time-varying parameter VAR (TVP-VAR) models, see Primiceri (2005), with the difference that we additionally need to draw the break processes $\boldsymbol{\kappa}_t = (\kappa_{bt}, \kappa_{st}, \kappa_{at})'$ and the break probabilities $\boldsymbol{\pi} = (\pi_b, \pi_s, \pi_a)'$.

The Gibbs sampler is split into four steps, which is iterated over until we have a sufficient number of posterior draws,

1. Draw the coefficients $\boldsymbol{B} = \{\boldsymbol{\beta}_t\}_{t=1}^T$ and their break process $\boldsymbol{\kappa}_b = (\kappa_{b,1}, \ldots, \kappa_{b,T})$.

2. Draw the volatilities $\boldsymbol{S} = \{\log \boldsymbol{\sigma}_t^2\}_{t=1}^T$ and their break process $\boldsymbol{\kappa}_s = (\kappa_{s,1}, \ldots, \kappa_{s,T})$.

3. Draw the covariance terms $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_T)$ and its break process $\boldsymbol{\kappa}_a = (\kappa_{a,1}, \ldots, \kappa_{a,T})$.

4. Draw the parameters in the state equations $\boldsymbol{\theta} = (\boldsymbol{\Phi}, \boldsymbol{Q}, \boldsymbol{\pi})$.

Each of the first three steps consists of writing (parts of) the VAR model into state space form such that we can draw the state variables in two steps,

a. Draw the break process $\boldsymbol{\kappa}_k$ from the full conditional posterior, with the state variables integrated out, using the algorithm of Gerlach et al. (2000).

b. Draw the state variables, i.e. the long-run mean and the time-varying part, using the simulation smoother of Durbin and Koopman (2002).

C.1 Step 1: Drawing the coefficients

Define $\boldsymbol{y}_t = (r_t, z_t)'$ and $\boldsymbol{x}_t = (1, z_{t-1})'$, the left hand side and right hand side variables, and $\tilde{\boldsymbol{\beta}}_t = \boldsymbol{\beta}_t - \boldsymbol{\mu}_b$, the coefficients' deviation from the mean at time t. Then, we can write

the state space model,

$$\boldsymbol{y}_t = (\boldsymbol{I}_n \otimes \boldsymbol{x}_t)(\boldsymbol{\mu}_b + \tilde{\boldsymbol{\beta}}_t) + \boldsymbol{\varepsilon}_t, \tag{C.1}$$

$$\tilde{\boldsymbol{\beta}}_t = f(\kappa_{bt}, \boldsymbol{\Phi}_b)\tilde{\boldsymbol{\beta}}_{t-1} + \kappa_{bt}\boldsymbol{\eta}_t, \tag{C.2}$$

$$\boldsymbol{\mu}_b = \boldsymbol{\mu}_b, \tag{C.3}$$

where $\boldsymbol{\varepsilon}_t \sim \mathrm{N}(\boldsymbol{0}, \boldsymbol{\Omega}_t)$ and $\boldsymbol{\eta}_t \sim \mathrm{N}(\boldsymbol{0}, \boldsymbol{Q}_b)$ are independent, $\otimes$ is the Kronecker product and $\boldsymbol{I}_n$ the $n \times n$ identity matrix, with $n = 2$ the number of left hand side variables.

Now, we can first apply the algorithm of Gerlach et al. (2000) to efficiently draw the break process $\boldsymbol{\kappa}_b$ from the full conditional posterior where the state variables have been integrated out. Then, conditional on the break process $\boldsymbol{\kappa}_b$, we use the simulation smoother of Durbin and Koopman (2002) to draw the state variables $\tilde{\boldsymbol{\beta}}_t$ and $\boldsymbol{\mu}_b$.

C.2 Step 2: Drawing the volatilities

Define

$$\boldsymbol{\varepsilon}_t^* = \boldsymbol{A}_t\hat{\boldsymbol{\varepsilon}}_t = \boldsymbol{A}_t\big(\boldsymbol{y}_t - (\boldsymbol{I}_n \otimes \boldsymbol{x}_t)\boldsymbol{\beta}_t\big) = \begin{bmatrix} 1 & 0 \\ \alpha_t & 1 \end{bmatrix} \begin{pmatrix} \hat{\varepsilon}_{1t} \\ \hat{\varepsilon}_{2t} \end{pmatrix} = \begin{pmatrix} \hat{\varepsilon}_{1t} \\ \alpha_t\hat{\varepsilon}_{1t} + \hat{\varepsilon}_{2t} \end{pmatrix}. \tag{C.4}$$

Next, transform $\boldsymbol{\varepsilon}_t^*$ into $\boldsymbol{\varepsilon}_t^{**} = \log\big((\boldsymbol{\varepsilon}_t^*)^2 + \bar{c}\big)$, with $\bar{c} = 0.0001$ the off-set constant to avoid numerical issues.Further, define $\widetilde{\log \boldsymbol{\sigma}_t^2} = \log \boldsymbol{\sigma}_t^2 - \boldsymbol{\mu}_s$, the volatilities' deviation from the mean at time t. Then, we can write the state space model,

$$\boldsymbol{\varepsilon}_t^{**} = \boldsymbol{\mu}_s + \widetilde{\log \boldsymbol{\sigma}_t^2} + \boldsymbol{e}_t, \tag{C.5}$$

$$\widetilde{\log \boldsymbol{\sigma}_t^2} = f(\kappa_{st}, \boldsymbol{\Phi}_s)\widetilde{\log \boldsymbol{\sigma}_{t-1}^2} + \kappa_{st}\boldsymbol{\zeta}_t, \tag{C.6}$$

$$\boldsymbol{\mu}_s = \boldsymbol{\mu}_s, \tag{C.7}$$

where $\boldsymbol{\zeta}_t \sim \mathrm{N}(\boldsymbol{0}, \boldsymbol{Q}_s)$ independent from $\boldsymbol{e}_t$.

The state space model is non-Gaussian, because the disturbances $\boldsymbol{e}_t$ follow a χ^2 distribution with one degree of freedom. Carter and Kohn (1997) and Kim et al. (1998)

show that it can be accurately approximated using a mixture of normals. We use the mixture of seven normals used by Kim et al. (1998), and can then consecutively draw the break process $\boldsymbol{\kappa}_s$, the time-varying part of $\widetilde{\log \boldsymbol{\sigma}^2}$ and the long-run mean $\boldsymbol{\mu}_s$.

Stroud et al. (2011) suggest a Metropolis-Hastings step to correct for the approximation error of using a mixture of normals. Kim et al. (1998) show that the approximation error is negligible, though, and Del Negro and Primiceri (2015) confirm this with a TVP-VAR model. Therefore, we also exclude a Metropolis-Hastings step.

C.3 Step 3: Drawing the covariance term

Define $\hat{\boldsymbol{\varepsilon}}_t = \boldsymbol{y}_t - (\boldsymbol{I}_n \otimes \boldsymbol{x}_t)\boldsymbol{\beta}_t$, and $\boldsymbol{A}_t\hat{\boldsymbol{\varepsilon}}_t = \boldsymbol{u}_t$, with $\boldsymbol{u}_t \sim \mathrm{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_t\boldsymbol{\Sigma}_t')$, then we can use the lower triangular structure of $\boldsymbol{A}_t$ to rewrite $\hat{\boldsymbol{\varepsilon}}_t$ as

$$\hat{\boldsymbol{\varepsilon}}_t = \boldsymbol{C}_t\alpha_t + \boldsymbol{u}_t = \begin{pmatrix} 0 \\ -\hat{\varepsilon}_{1t} \end{pmatrix} \alpha_t + \boldsymbol{u}_t. \tag{C.8}$$

Further, define $\tilde{\alpha}_t = \alpha_t - \mu_a$ the covariance term's deviation from the long run mean. Then, we can write the state space model,

$$\hat{\varepsilon}_{2t} = -\hat{\varepsilon}_{1t}(\mu_a + \tilde{\alpha}_t) + u_{2t}, \tag{C.9}$$

$$\tilde{\alpha}_t = f(\kappa_{at}, \phi_a)\tilde{\alpha}_{t-1} + \kappa_{at}\xi_t, \tag{C.10}$$

$$\mu_a = \mu_a, \tag{C.11}$$

where $u_{2t} \sim \mathrm{N}(0, \sigma_{2t}^2)$ and $\xi_t \sim \mathrm{N}(0, q_a^2)$ are independent.

We again apply steps (a) and (b) to draw the break process $\boldsymbol{\kappa}_a$, followed by the time-varying part of the covariance term $\tilde{\boldsymbol{\alpha}} = (\tilde{\alpha}_1, \ldots, \tilde{\alpha_T})$ and the long-run mean μ_a.

C.4 Step 4: Drawing the state equation parameters

The full conditional posterior distribution for the diagonal elements of the autoregressive parameter $\boldsymbol{\Phi}_k$ is a truncated normal distribution,

$$\text{diag}(\boldsymbol{\Phi}_k)|\ldots \sim \text{N}(\overline{\boldsymbol{m}}_k, \overline{\boldsymbol{V}}_k) I(|\phi_{ki}| < 1, \forall i = 1, \ldots, n_k), \tag{C.12}$$

where $I(A)$ is one if A holds and zero otherwise, n_k the number of diagonal elements in $\boldsymbol{\Phi}_k$, with variance

$$\overline{\boldsymbol{V}}_b = \left(\underline{\boldsymbol{V}}_b^{-1} + \sum_{t=2}^{T} \kappa_{bt} (\tilde{\boldsymbol{\beta}}_{t-1}' \boldsymbol{Q}_b^{-1} \tilde{\boldsymbol{\beta}}_{t-1}) \right)^{-1}, \tag{C.13}$$

$$\overline{\boldsymbol{V}}_s = \left(\underline{\boldsymbol{V}}_s^{-1} + \sum_{t=2}^{T} \kappa_{st} \big((\widetilde{\log \boldsymbol{\sigma}_{t-1}^2})' \boldsymbol{Q}_s^{-1} \widetilde{\log \boldsymbol{\sigma}_{t-1}^2} \big) \right)^{-1}, \tag{C.14}$$

$$\overline{v}_a^2 = \left(\underline{v}_a^{-2} + q_a^{-2} \sum_{t=2}^{T} \kappa_{at} \tilde{\alpha}_{t-1}^2 \right)^{-1}, \tag{C.15}$$

and mean

$$\overline{\boldsymbol{m}}_b = \overline{\boldsymbol{V}}_b \left(\underline{\boldsymbol{V}}_b^{-1} \underline{\boldsymbol{m}}_b + \sum_{t=2}^{T} \kappa_{bt} (\tilde{\boldsymbol{\beta}}_{t-1}' \boldsymbol{Q}_b^{-1} \tilde{\boldsymbol{\beta}}_t) \right), \tag{C.16}$$

$$\overline{\boldsymbol{m}}_s = \overline{\boldsymbol{V}}_s \left(\underline{\boldsymbol{V}}_s^{-1} \underline{\boldsymbol{m}}_s + \sum_{t=2}^{T} \kappa_{st} \big((\widetilde{\log \boldsymbol{\sigma}_{t-1}^2})' \boldsymbol{Q}_s^{-1} \widetilde{\log \boldsymbol{\sigma}_t^2} \big) \right), \tag{C.17}$$

$$\overline{m}_a = \overline{v}_a^2 \left(\underline{v}_a^{-2} \underline{m}_a + q_a^{-2} \sum_{t=2}^{T} \kappa_{at} (\tilde{\alpha}_{t-1} \tilde{\alpha}_t) \right), \tag{C.18}$$

and the off-diagonal elements of $\boldsymbol{\Phi}_k$ are zero.

We draw from the truncated normal distribution using an accept-reject algorithm with a normal distribution with the same mean and variance to reduce computational costs. If after a 1,000 tries no draw has been accepted, we draw directly from the truncated normal distribution.

The full conditional posterior distribution for the break probability π_k is a beta

distribution,

$$\pi_k|\ldots \sim \text{Beta}(\overline{a}_k, \overline{b}_k), \tag{C.19}$$

where

$$\overline{a}_k = \underline{a}_k + \sum_{t=2}^{T} \kappa_{kt}, \tag{C.20}$$

$$\overline{b}_k = \underline{b}_k + \sum_{t=2}^{T} (1 - \kappa_{kt}), \tag{C.21}$$

for $k \in \{b, s, a\}$.

The full conditional posterior distribution of the break size $\boldsymbol{Q}_k$ is an inverse Wishart distribution,

$$\boldsymbol{Q}_k|\ldots \sim \text{IW}\big(\overline{\boldsymbol{W}}_k, \overline{\nu}_k\big), \tag{C.22}$$

with location parameter

$$\overline{\boldsymbol{W}}_b = \underline{\boldsymbol{W}}_b \underline{\nu}_b + \sum_{t=2}^{T} \kappa_{bt} \big(\tilde{\boldsymbol{\beta}}_t - f(\kappa_{bt}, \boldsymbol{\Phi}_b)\tilde{\boldsymbol{\beta}}_{t-1}\big)\big(\tilde{\boldsymbol{\beta}}_t - f(\kappa_{bt}, \boldsymbol{\Phi}_b)\tilde{\boldsymbol{\beta}}_{t-1}\big)', \tag{C.23}$$

$$\overline{\boldsymbol{W}}_s = \underline{\boldsymbol{W}}_s \underline{\nu}_s + \sum_{t=2}^{T} \kappa_{st} \left(\widetilde{\log \boldsymbol{\sigma}_t^2} - f(\kappa_{st}, \boldsymbol{\Phi}_s)\widetilde{\log \boldsymbol{\sigma}_{t-1}^2}\right)\left(\widetilde{\log \boldsymbol{\sigma}_t^2} - f(\kappa_{st}, \boldsymbol{\Phi}_s)\widetilde{\log \boldsymbol{\sigma}_{t-1}^2}\right)', \tag{C.24}$$

$$\overline{w}_a = \underline{w}_a \underline{\nu}_a + \sum_{t=2}^{T} \kappa_{at} \big(\tilde{\alpha}_t - f(\kappa_{at}, \phi_a)\tilde{\alpha}_{t-1}\big)^2, \tag{C.25}$$

and degrees of freedom

$$\overline{\nu}_k = \underline{\nu}_k + \sum_{t=2}^{T} \kappa_{kt}, \tag{C.26}$$

for $k \in \{b, s, a\}$.

Appendix D MCMC convergence analysis

We follow the convergence analysis by Groen et al. (2013) in their appendix. That is, we compute inefficiency factors and analyze the convergence of the MCMC chain using the Geweke (1992) test, for our baseline model, the MI model with $\mathrm{E}[\pi_b] = 1/60$.

The inefficiency factors in table Table D.1 are calculated as $\mathrm{IF} = 1 + 2\sum_{i=1}^{\infty} \rho_i$, with ρ_i the i-th order autocorrelation of the posterior draws of a parameter. The autocorrelation is computed using the Newey and West (1987) estimator with a Bartlett kernel and 4% bandwidth. A guideline is that you need approximately 100 times the inefficiency factor such that at most 1% of the variation is due to the data, see Kim et al. (1998) and Groen et al. (2013) (appendix). Most parameters are reasonably converged for 10,000 draws, the number of retained draws we use. The results do suggest that we require extra draws to get a more accurate estimate of the parameter in the covariance term's state equation. The inefficiency factor for the long run mean is small, so it only concerns the parameters describing the dynamics of α_t.

The Geweke (1992) tests whether the first 20% and the last 40% of the draws have an equal mean. The rejection rates in table Table D.2 show that in general the MCMC chain is converged. However, the null hypothesis is strongly rejected for the parameters governing the covariance's break process. We experimented with the burn-in, and test results are mostly due to sample selection rather than a lack of convergence. Due to the high autocorrelation, it depends on the selected subsamples whether the null of equal means is rejected.

Combining the inefficiency factors and Geweke (1992) test results, only the parameters in the covariance term's state equation are not well identified. It is not surprising, because there is little instability found in this the covariance term. We are not too concerned about the effect on the overall convergence, given that the test results for the other parameters are fine. Concluding, we find that the overall convergence of the sampler for the current number of draws and burn-in is satisfactory.

Table D.1: Inefficiency factors

Parameter	Number	Median	Mean	Min	Max	5% quantile	95% quantile
			Panel A: Intercepts and loadings				
$\tilde{\boldsymbol{\beta}}$	3,360	15.44	16.92	2.60	49.39	5.06	37.62
$\boldsymbol{\mu}_b$	4	52.25	52.78	48.18	58.46		
$\boldsymbol{\kappa}_b$	840	19.00	20.72	8.01	65.09	16.05	30.33
π_b	1	64.06	64.06	64.06	64.06		
$\boldsymbol{\Phi}_b$	4	95.07	101.69	51.97	164.66		
$\boldsymbol{Q}_b$	16	53.85	53.63	51.22	55.46		
			Panel B: Variances				
$\widetilde{\log \boldsymbol{\sigma}^2}$	1,680	5.23	9.89	1.61	83.15	2.69	27.90
$\boldsymbol{\mu}_s$	2	16.06	16.06	6.66	25.46		
$\boldsymbol{\kappa}_s$	840	1.23	1.17	0.52	5.24	0.75	1.84
$\boldsymbol{\pi}_s$	1	20.49	20.49	20.49	20.49		
$\boldsymbol{\Phi}_s$	2	96.47	96.47	28.74	164.19		
$\boldsymbol{Q}_s$	4	61.21	62.62	42.29	85.77		
			Panel C: Covariance term				
$\boldsymbol{\alpha}$	840	1.35	1.44	0.72	6.04	0.92	2.24
μ_a	1	6.12	6.12	6.12	6.12		
$\boldsymbol{\kappa}_a$	840	81.85	81.85	74.23	88.31	77.79	85.93
π_a	1	289.44	289.44	289.44	289.44		
ϕ_a	1	95.08	95.08	95.08	95.08		
q_a	1	203.38	203.38	203.38	203.38		
$\log p(\boldsymbol{Y}\|\boldsymbol{\kappa}_b)$	1	4.40	4.40	4.40	4.40		

This table presents a summary of the inefficiency factors for the parameters of the mixture innovation model, where the break probability is unknown, and the hyperparameters are set as $\underline{a}_b = 1$,$\underline{b}_b = 59$, and $\underline{a}_s = \underline{b}_s = \underline{a}_a = \underline{b}_a = 1$. The inefficiency factors are estimated using Newey and West (1987) estimator.

Table D.2: Geweke test results

Parameter	Number	10% rejection rate	5% rejection rate	1% rejection rate
Panel A: Intercept and loadings				
$\tilde{\boldsymbol{\beta}}$	3,360	0.001	0.000	0.000
$\boldsymbol{\mu}_b$	4	0.000	0.000	0.000
$\boldsymbol{\kappa}_b$	840	0.004	0.000	0.000
π_b	1	0.000	0.000	0.000
$\boldsymbol{\Phi}_b$	4	0.000	0.000	0.000
$\boldsymbol{Q}_b$	16	0.000	0.000	0.000
Panel B: Variances				
$\widetilde{\log \boldsymbol{\sigma}^2}$	1,680	0.080	0.039	0.007
$\boldsymbol{\mu}_s$	2	0.000	0.000	0.000
$\boldsymbol{\kappa}_s$	840	0.092	0.061	0.010
$\boldsymbol{\pi}_s$	1	0.000	0.000	0.000
$\boldsymbol{\Phi}_s$	2	0.000	0.000	0.000
$\boldsymbol{Q}_s$	4	0.000	0.000	0.000
Panel C: Covariance term				
$\boldsymbol{\alpha}$	840	0.096	0.056	0.018
μ_a	1	0.000	0.000	0.000
$\boldsymbol{\kappa}_a$	840	1.000	1.000	1.000
π_a	1	1.000	1.000	1.000
ϕ_a	1	1.000	1.000	1.000
q_a	1	1.000	1.000	1.000
$\log p(\boldsymbol{Y}\|\boldsymbol{\kappa}_b)$	1	0.000	0.000	0.000

This table presents the rejection rates of the Geweke (1992) test, testing whether the first 20% and the last 40% of the draws have an equal mean, for the parameters of the mixture innovation model, where the break probability is unknown, and the hyperparameters are set as $\underline{a}_b = 1$,$\underline{b}_b = 59$, and $\underline{a}_s = \underline{b}_s = \underline{a}_a = \underline{b}_a = 1$.

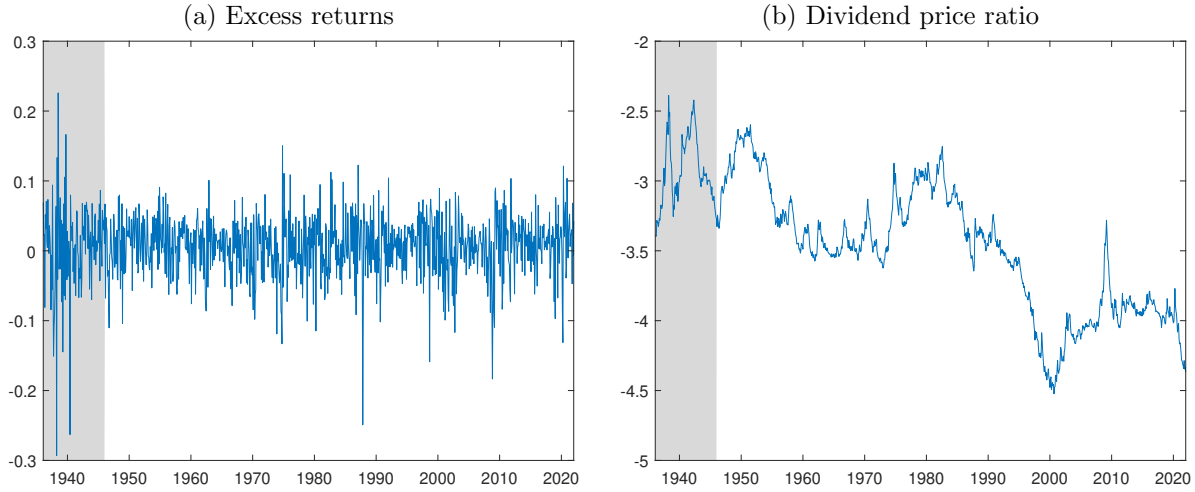
Appendix E Data

Table E.1: Summary statistics

	Excess market return	Dividend price ratio
Mean	0.006	−3.514
Median	0.010	−3.479
Standard deviation	0.042	0.443
Skewness	−0.655	−0.112
Kurtosis	5.207	2.215
AR(1)	0.042	0.995
AR(12)	0.043	0.932
ARCH test (p-value)	0.000	0.000

The table presents summary statistics of the excess log return and log dividend price ratio for the period January 1946 – December 2021. AR(x) is the x-th order autocorrelation. Engle's (1982) ARCH test is based on residuals from a static VAR(1) model, estimated using maximum likelihood.

Figure E.1: Time series of data



The figures present the monthly time series of the excess log returns (a) and the log dividend price ratio (b) for the period January 1936 – December 2021, where the shaded area (January 1936 – December 1945) is used for prior calibration.