FEDERAL RESERVE BANK *of* ATLANTA WORKING PAPER SERIES

Bayesian Inference and Prediction of a Multiple-Change-Point Panel Model with Nonparametric Priors

Mark Fisher and Mark J. Jensen

Working Paper 2018-2a
June 2018

Abstract: Change point models using hierarchical priors share in the information of each regime when estimating the parameter values of a regime. Because of this sharing, hierarchical priors have been very successful when estimating the parameter values of short-lived regimes and predicting the out-of-sample behavior of the regime parameters. However, the hierarchical priors have been parametric. Their parametric nature leads to global shrinkage that biases the estimates of the parameter coefficient of extraordinary regimes toward the value of the average regime. To overcome this shrinkage, we model the hierarchical prior nonparametrically by letting the hyperparameter's prior—in other words, the hyperprior—be unknown and modeling it with a Dirichlet processes prior. To apply a nonparametric hierarchical prior to the probability of a break occurring, we extend the change point model to a multiple-change-point panel model. The hierarchical prior then shares in the cross-sectional information of the break processes to estimate the transition probabilities. We apply our multiple-change-point panel model to a longitudinal data set of actively managed, U.S. equity, mutual fund returns to measure fund performance and investigate the chances of a skilled fund being skilled in the future.

JEL classification: C11, C14, C41, G11, G17

Key words: Bayesian nonparametric analysis, change points, Dirichlet process, hierarchical priors, mutual fund performance

https://doi.org/10.29338/wp2018-02

The authors thank Vikas Agarwal, Sylvia Frühwirth-Schnatter, and Paula Tkac for helping to make this paper better. They also thank the conference participants at the 2015 Bayesian Workshop of the Rimini Centre of Economic Analysis in Rimini, Italy; the 2015 European Seminar on Bayesian Econometrics in Gerzensee, Switzerland; the 2015 Conference on Computational and Financial Econometrics in London; the 2016 All Georgia Finance Conference in Atlanta, Georgia; the 2017 XVII ESTE Brazilian School of Time Series and Econometrics in São Carlos, Brazil; the department members of the Vienna University of Economics and Business, where one of the authors visited; McMasters University; the University of Montreal; Clemson University; and the University of North Carolina-Charlotte for their helpful comments and suggestions. The views expressed here are the authors' and not necessarily those of the Federal Reserve Bank of Atlanta or the Federal Reserve System. Any remaining errors are the authors' responsibility.

Please address questions regarding content to Mark Fisher, Research Department, Federal Reserve Bank of Atlanta, 1000 Peachtree Street NE, Atlanta, GA 30309-4470, 404-498-8757, mark@markfisher.net, or Mark J. Jensen, Research Department, Federal Reserve Bank of Atlanta, 1000 Peachtree Street NE, Atlanta, GA 30309-4470, 404-498-8019, mark.jensen@atl.frb.org.

1 Introduction

In economics and finance there are many instances where longitudinal data changes regimes at different points in time when experiencing structural breaks. For example, models of the real gross domestic products of the world's economies, or the returns from the group of actively managed mutual funds, can have parameters that change over time and across the cross-section. When measuring the performance of actively managed mutual funds it is the fund-regime performance parameter, alpha, that changes from regime to regime and from fund to fund. Allowing for changes in each fund's alpha at different points in time allows there to be periods when some funds are skilled and other times when the same funds are unskilled. To model this regime parameter behavior over a cross-section of individuals our first contribution is to extend the change point model to a panel of change point processes.[1] In particular, we model the performance of actively managed mutual funds with a multiple-change-point panel model where the skill level of each fund follows its own change-point process.

Hierarchical priors have been an important advancement to the estimation of the regime parameters of structural break models (see Pesaran et al. (2006), Koop & Potter (2007), Geweke & Jiang (2011), Maheu & Song (2014) and Song (2014)). Under hierarchical priors the hyperparameters are unknown and their values are learned about through the information in the regimes. By mixing the conditional hierarchical prior over the posterior of its hyperparameters, the hierarchical prior's posterior predictive distribution flexibly models how the parameters are distributed over the regimes.

These hierarchical priors, however, are restrictive in the sense that the prior for the hyperparameter, the hyperprior distribution, is assumed to be known. Assuming the hyperprior distribution is known can lead to the undesirable outcome of the regime parameter estimates shrinking towards the global average of the regimes (see Gelman et al. (2013), Chapter 5). For example, when measuring mutual fund skill the performance estimate for a highly skilled mutual fund shrinks to that of an ordinary fund (see Jones & Shanken (2005)).

To provide a better and more robust estimate of the parameters we model the hierarchical prior nonparametrically by letting the hyperprior be unknown, giving it a Dirichlet process prior, and estimating it along with the hyperparameters. Nonparametric hierarchical priors span the space of distributions, including multimodal, skewed, and kurtotic

[1] Frühwirth-Schnatter & Kaufmann (2006) model a panel of bank lending series with a multiple-structural-break model. Billio et al. (2016) also propose a panel of Markov switching vector autoregressive models but their panel is relatively small compared to the size of the panel we envision here.

distributions (see Lijoi et al. (2005)). This spanning property is especially important when estimating the parameters of an extraordinary individual or for an extreme regime. Our second contribution relaxes the parametric nature of the hierarchical priors and flexibly models the hierarchical priors nonparametrically by letting the hyperprior be unknown.

The last contribution of the paper is found by applying our multiple-change-point panel model to a cross-section of mutual fund return data and estimating the alphas for each fund and their different regimes. Estimates of these fund-regime alphas help answer if above market returns by a fund today are any indication of future excellence by the fund.[2] The answer depends on a number of factors, for instance, who the fund manager is, whether the fund changes its strategy or objective, if the fund adjusts its risk exposure, what the fund's flow of assets under management is, and how the fund manager is compensated. A change in any one of these factors can result in a new regime and a different value for alpha. We find there to be overwhelming empirical evidence in favor of mutual fund performance following a change-point process so there are no guarantees future performance will mirror today's.

Our plan for the paper is to describe in Section 2 our multiple-change-point panel model. Section 3 then defines the nonparametric hierarchical priors for the model parameters. In that section we also define the Dirichlet process prior for an unknown hyperprior distribution. A Markov chain Monte Carlo (MCMC) sampler of the unknowns is then described in Section 4. Section 4.4 explains how the posterior draws from the MCMC sampler are used to infer how the unknowns are distributed over mutual funds and their regimes. In Section 5 we apply the multiple-change-point panel model, using both nonparametric and parametric, hierarchical, priors, to a unbalanced panel of actively managed, US equity, mutual fund return data. Section 6 concludes with a summary of our findings.

2 Modeling a panel of multiple-change-point processes

Our multiple-change-point panel model extends the models of Chib (1998), Pesaran et al. (2006), Koop & Potter (2007) and Geweke & Jiang (2011). Instead of there being a single change point process, our model consists of a cross-section of multiple-change-point processes. Given our interest in measuring the empirical performance of mutual funds and their ability to sustain above market returns we describe the model in terms of mutual fund returns.

[2]See Grinblatt & Sheridan (1992), Hendricks et al. (1993), Goetzmann & Ibbotson (1994), Elton et al. (1996), Bollen & Busse (2004), and Busse & Irvine (2006) for empirical evidence favoring a mutual fund performance today predicting future performance and Carhart (1997) for a contrary view.

Let $y_{i,t}$, $i = 1, \ldots, J$, and, $t = \tau_i, \ldots, T_i$, where $1 \leq \tau_i$, be the risk-free adjusted, gross returns in month t for the ith mutual fund from a longitudinal data set where the length of the series is $\mathcal{T}_i = T_i - \tau_i + 1$. In contrast to the existing literature on change-point models, the lengths of the time series are all small relative to the size of the cross-section, J. In addition, the panel does not have to be balanced. As a result our model can handle return histories, $Y_{i,T_i} = (y_{i,\tau_i}, \ldots, y_{i,T_i})'$, that are of different lengths; ie., $\mathcal{T}_i$ does not have to equal $\mathcal{T}_{i'}$. The model can also handle return histories that do not line up in time so τ_i does not need to be equal to $\tau_{i'}$.

Following Chib (1998) we model the change point processes with a hidden Markov chain process. Let M_i be the maximum number of regimes the ith mutual fund can experience. In addition, let $m \in \{1, \ldots, M_i\}$ index the fund-regimes where the probability of switching to a new fund-regime is q_i. Fund i's restricted Markov break process at observation t is denoted by $s_{i,t}$, where $t = \tau_i, \tau_i + 1, \ldots$. The change point process can then be represented by the following hierarchical prior

$$Pr(s_{i,t+1} = l | s_{i,t} = m, q_i) = \begin{cases} 1 - q_i & l = m < M_i, \\ q_i & l = m + 1, \\ 1 & l = m = M_i, \\ 0 & \text{otherwise}, \end{cases} \tag{1}$$

$Pr(s_{i,\tau_i} = 1) = 1$, and where the next level in the hierarchy is the prior for q_i.

According to Eq. (1) once a break occurs a fund cannot return to an earlier regime, nor can it skip over a regime. Such change point behavior is not restrictive when measuring the persistence in mutual fund performance. Mutual fund skill and stock picking strategies are difficult to replicate over different time periods. Hence, even if a past trading strategy could be adopted, or a known skilled manager were to be hired, performance today would most likely differ from the past given the additional information available today and the changing market environment.

In Eq. (1), once $s_{i,t}$ reaches the M_ith regime it stays there and cannot experience another break. To avoid a pile-up on M_i and the problems found in Koop & Potter (2007), we follow Bauwens et al. (2015) and set M_i apriori and monitor $s_{i,t}$ to ensure that it never reaches M_i before t equals T_i. In addition, to speed up estimation we assign smaller M_is to those funds with shorter histories, $\mathcal{T}_i$ (see Fig. 1 for a graph of the value for the M_is). This reduces the need to estimate a large number of out-of-sample, fund-regime parameters.[3]

[3]Our approach requires estimating the parameter values for all M_i regimes and evaluating their likelihoods. It would be computationally less demanding if we could use the approach of Maheu & Song (2014) and let the number of change points be unknown. However, this would require the priors for the regime parameters to be joint conjugate which would restrict the flexibility of our nonparametric hierarchical priors.

The transition probability, q_i, is the same as in Chib (1998) and Pesaran et al. (2006), in that it is constant over time. Notice also that q_i does not depend on the regime m (see McCulloch & Tsay (1993) and Geweke & Jiang (2011) for models with the same assumption). Each fund's transition probability is different but we assume they all come from the same cross-sectional distribution, $q_i \overset{iid}{\sim} \pi_q$, $i = 1, \ldots, J$.

As pointed out by Koop & Potter (2007), letting q_i be constant has consequences. It means the duration of a regime follows a Geometric distribution where shorter length regimes have a higher probability of occurrence than do longer regimes. Although restrictive, empirically we find that by implicitly modeling the duration distribution as an unknown distribution through an unknown π_q, the probability of a long-lived regime occurring is not that much lower than the probability of a short-lived regime.

To simplify our exposition on estimating our multiple-change-point panel model, we limit the regime parameters to the skill level and the variance of the mutual fund returns in a zero risk-factor model. Later, in our empirical analysis, we let the returns follow a four factor risk model where the unknown beta coefficients follow the same change point process underlying the regime parameter values of skill and variance.

Let α_{im} and σ^2_{im} be the level of skill and variance of the mth regime for the ith fund. Using a dynamic structural investment model of mutual fund performance, Koijen (2014) shows that changes in a fund's strategy or its manager is equivalent to α_{im} changing value with each regime m. When $s_{i,t} = m$, the conditional sampling distribution of $y_{i,t}$ is then defined to be

$$p(y_{i,t}|Y_{i,t-1}, s_{i,t} = m) \equiv N(\alpha_{im}, \sigma^2_{im}). \tag{2}$$

We could replace normally distributed returns with the scaled normal mixture representation of a Student-t distribution. A time-varying stochastic volatility process for the variances could also be used in place of the constant fund-regime variance. However, monthly mutual fund returns do not exhibit the liptokurtosis and time-varying volatility prevalent in daily stock return data. Instead, monthly mutual fund returns are known for their homoskedasticity and normality.

Let $\mathcal{Y} := \{Y_{i,T_i}\}_{i=1}^J$ be the entire panel of mutual fund return data, and $\boldsymbol{\alpha}_i = (\alpha_{i1}, \ldots, \alpha_{iM_i})'$, and, $\boldsymbol{\sigma}^2_i = (\sigma^2_{i1}, \ldots, \sigma^2_{iM_i})'$, be all the regime parameters for the ith mutual fund. Denote the history of the latent change points up to time t as $S_{it} = (s_{i,\tau_i}, \ldots, s_{i,t})$ so that S_{iT_i} contains the entire in-sample history of the change points. If we assume trading strategies and stock picking skills are proprietary secrets, and, therefore, cannot be copied by, or transferred to,

another fund, the joint likelihood is written as

$$p(\mathcal{Y}|\boldsymbol{\alpha}, \boldsymbol{\sigma}^2, \mathcal{S}) = \prod_{i=1}^{J} \prod_{t=\tau_i}^{T_i} p(y_{i,t}|\alpha_{is_{i,t}}, \sigma^2_{is_{i,t}}), \tag{3}$$

where $\boldsymbol{\alpha} := \{\boldsymbol{\alpha}_i\}_{i=1}^J$, $\boldsymbol{\sigma}^2 := \{\boldsymbol{\sigma}_i^2\}_{i=1}^J$, and $\mathcal{S} := \{S_{iT_i}\}_{i=1}^J$. This factoring of the likelihood simplifies both the exposition and the computation, but it is not necessary. One could choose to follow the approach of Jones & Shanken (2005) and model the cross-correlation between fund returns with a latent residual factor structure.

3 Nonparametric hierarchical priors

What one assumes about the priors for the regime parameters and transition probabilities plays an important role in the behavior of the multiple-change-point panel model and the measurement of mutual fund performance and its persistence. Specification of the prior can affect the expectation about the future performance of the actively managed mutual fund industry, the level of certainty in a fund's measure of performance, and the inference about how skill, risk taking, and the probability of a new regime, are distributed across the population of mutual funds and their regimes.

To answer these and other mutual fund performance related questions with our multiple-change-point panel model we denote the prior distribution for α_{im}, σ^2_{im}, and, q_i, as π_α, π_σ, and π_q, respectively. We also assume each prior is unknown and modeled with a nonparametric hierarchical prior where the hyperprior is unknown and follows the Dirichlet process (DP) of Ferguson (1973).

Our approach is similar but slightly different from Frühwirth-Schnatter & Kaufmann's (2008) unobserved heterogeneous, model-based, clustering approach. The two approaches are similar in that both model the hyperprior as an unknown. But, they differ in that Frühwirth-Schnatter & Kaufmann (2008) models the hyperprior with a finite mixture, whereas our hyperprior is almost surely an infinite ordered mixture distribution.[4]

Formally, our nonparametric hierarchical prior for the panel of skill measurements, α_{im}, is defined as

$$\alpha_{im}|a_{\alpha,im}, h^2_{\alpha,im} \sim N(a_{\alpha,im}, h^2_{\alpha,im}), \tag{4}$$

$$(a_{\alpha,im}, h^2_{\alpha,im}) \sim G_\alpha, \tag{5}$$

[4]Chan & Koop (2014) also use Frühwirth-Schnatter & Kaufmann's (2008) unobserved heterogeneous, model-based, clustering approach to group together structural break series.

where G_α is a unknown hyperprior distribution. A similar conditional prior representation exists for our nonparametric hierarchical prior of the multiple-change-point variances, σ^2_{im}, where

$$\sigma^2_{im}|h^2_{\sigma,im} \sim \text{Inv-Gamma}(\nu_\sigma/2, \nu_\sigma h^2_{\sigma,im}/2), \tag{6}$$

$$h^2_{\sigma,im} \sim G_\sigma. \tag{7}$$

Note the shape hyperparameter, ν_σ, is assumed to be known and is equal to the same value for all J mutual funds. In Eq. (6) and (7), $h^2_{\sigma,im}$ is the unknown scale hyperparameter whose distribution, G_σ, is unknown.[5]

We define the nonparametric hierarchical prior for the transition probabilities, q_i, to be

$$q_i|j_i, k_i \sim \text{Beta}(j_i, k_i - j_i + 1), \tag{8}$$

$$(j_i, k_i) \sim G_q, \tag{9}$$

where $k_i \in \mathbb{N}$, $j_i \in \{1, \ldots, k_i\}$, and G_q is the unknown hyperprior distribution. Unlike the regime parameters' hierarchical priors, which model how the parameters are distributed across both fund and their regimes, Eq. (8)-(9) only models how the q_is are distributed across the J funds.[6]

Mixing the beta distribution in Eq. (8) over its hyperparameters leads to the flexible Bernstein prior. Research by Petrone & Wasserman (2002) shows the posterior of the Bernstein prior converging to the true data generating density of a continuous and bounded distribution residing on the unit interval. To our knowledge a hierarchical beta distribution has never been proposed before as a prior for an unknown parameter defined on the unit interval (see Fisher (2017)).

Hierarchical priors like those above are desirable since they facilitate the sharing of information across the cross-section of funds and their regimes through the priors' unknown hyperparameters, $a_{\alpha,im}$, $h^2_{\alpha,im}$, $h^2_{\sigma,im}$, j_i and k_i. For instance, the observed performance of a highly skilled manager or a particular trading strategy gets reflected in the posterior of the performance prior's hyperparameters $a_{\alpha,im}$ and $h^2_{\alpha,im}$. These posterior hyperparameter estimates then help reduce the uncertainty around how another fund might perform if it were to hire away the manager or adopt the same trading strategy. This sharing of information

[5] Because neither the conditional variances, nor their log transform, are normally distributed, our multiple-change-point panel model does not have the mixture innovation representation of Giordani & Kohn (2008). Therefore, we cannot sample the change point process with Gerlach et al.'s (2000) fast and efficient algorithm.

[6] If in the future we were interested in investigating the persistence of a particular regime, m_i, for instance, the skill of a highly successful manager, there is nothing in our approach that would prohibit it from being extended to the regime specific transition probability case.

through the learning of the hyperparameters is not possible when the parameters come from a prior distribution whose hyperparameters are fixed apriori.

Parametric hierarchical priors where the hyperprior distribution is assumed to be known have been used before to infer the parameters of change points and structural break models (see Pesaran et al. (2006), Koop & Potter (2007), Geweke & Jiang (2011) and Maheu & Song (2014)). In the context of mutual fund performance, assuming the hyperprior is known amounts to making a strong assumption about how the change point parameters are distributed across the funds and their regimes. For instance, assuming a normal, inverse-gamma hyperprior for G_α explicitly leads to the alphas being distributed over the population of mutual funds and their regimes in a unimodal fashion whose mode is the global average skill level. Given the abundant information contained in the cross-section of a longitudinal data set, this parametric assumption causes the posterior estimates to globally shrink towards the average fund-regime level. Alphas for the highly skilled and unskilled fund-regimes will end up looking ordinary like that of the skill level for a typical fund-regime. To overcome this global shrinkage we let the hyperpriors be unknown and, like the hyperparameters, learn about the hyperprior distributions by modeling them with DP priors.

3.1 Dirichlet process prior

The Dirichlet process has received considerable attention from the Bayesian nonparametric community as a prior to the mixture weights and location parameters of a infinite mixture representation of a unknown distribution (see Escobar & West (1995) and Hjort et al. (2010), and the chapters therein, for a good introduction to the DP). DPs have also been used extensively in economics and finance to model unknown distributions (see Chib & Tiwari (1988), Chib & Hamilton (2002), Hirano (2002), Jensen & Maheu (2010), and Bassetti et al. (2014)). In addition, Song (2014), Dufays (2015), Jochmann (2015) and Jin & Maheu (2016) apply the DP to structural break models to allow for an infinite number of possible regimes.

One reason for this attention is that the DP delivers a sparse mixture representation of unknown distributions. It is this sparsity that leads to the local shrinkage in the regime parameter estimates. Another reason is the DPs ease of use. As a conjugate distribution the DP lends itself to a straightforward sampler, one that quickly converges to draws from the posterior distribution of the nonparametric hierarchical prior. Under the DP prior our nonparametric hierarchical prior also nests the parametric hierarchical priors used by Chib (1998), Pesaran et al. (2006) and Geweke & Jiang (2011). These properties lead us to apply

the DP prior to the three hierarchical priors defined in Eq. (4)–(9).

Let the unknown hyperpriors have the following DP priors

$$G_\alpha \sim DP(\eta_\alpha, G_{\alpha,0}), \tag{10}$$

$$G_\sigma \sim DP(\eta_\sigma, G_{\sigma,0}), \tag{11}$$

$$G_q \sim DP(\eta_q, G_{q,0}). \tag{12}$$

The non-negative scalars, η_α, η_σ, and η_q, are the concentration parameters of the DPs. Their value is important since it determines the propensity of clustering and the variance in the DP; e.g., for the hyperprior G_σ, $\text{Var}[G_\sigma(A)] = [G_{\sigma,0}(A)(1 - G_{\sigma,0}(A))]/(1 + \eta_\sigma)$, where A is a measurable subset of $h^2_{\sigma,im}$ domain. A larger concentration parameter causes the DP to be less parsimonious and to have a lower propensity to cluster the hyperparameters together. For example, when $\eta_\sigma \to \infty$, each σ^2_{im} belongs to its own unique group.

To simplify our exposition we initially treat the concentration parameters as known values, but, empirically infer their value by sampling from their conditional posterior distribution. When sampled we assume a Log-Logistic$(1, 1)$ distribution for their prior.[7] Under this prior the probability of the DP having a second cluster, $u_\bullet \equiv \eta_\bullet/(1 + \eta_\bullet)$, is uniform over the unit interval. This probability goes to zero as $\eta_\bullet \to 0$, whereas it goes to one as $\eta_\bullet \to \infty$. We can then address the amount of empirical evidence in favor of a parametric hierarchical prior by measuring the posterior density of $u_\bullet$ at zero.

$G_{\sigma,0}$, $G_{\alpha,0}$, and $G_{q,0}$ are the base distributions of the DP priors and must be specified by the econometrician. These distributions equal the expected distribution of the DP prior; e.g., $G_{\sigma,0} \equiv E[G_\sigma]$. In our nonparametric hierarchical priors, the base distribution represents our initial view about how the hyperparameters are distributed over the relevant population. In this paper we choose base distributions that bring as little prior information about the population distribution to the analysis as possible.[8]

In our empirical application we set $G_{\alpha,0}$ equal to the conditional conjugate, normal, inverse-gamma distribution, NIG$(a_\alpha, h^2_\alpha | a_0, h^2_\alpha/\kappa_0, \nu_0/2, \nu_0 h^2_0/2)$, where the mean is $a_0 = 0$, the scaling of the variance is $\kappa_0 = 0.1$, the shape is $\nu_0/2 = 0.005$, and the scale is $\nu_0 h^2_0/2 = 0.005 \times 0.01$. With the NIG base distribution it follows from Eq. (4) that our initial guess for how the alphas are distributed over the funds and their regimes is represented by the

[7] The probability density function for the Log-Logistic$(1, 1)$ prior distribution of $\eta_\bullet$ is $1/(1 + \eta_\bullet)^2$ when $\eta_\bullet > 0$, zero otherwise.

[8] In subsequent research we have found that the empirical results in Section 5 are robust to the choice of the base distribution, as long as there is sufficient variation in the prior distributions drawn from the nonparametric hierarchical prior.

prior predictive Student-t distribution

$$\begin{aligned} E_{G_{\alpha,0}}\left[E_{G_\alpha}[\pi_\alpha(\alpha|G_\alpha)]\right] &= \int N(\alpha|a_\alpha, h_\alpha^2)\; NIG\left(a_\alpha, h_\alpha^2 \left| a_0, h_\alpha^2/\kappa_0, \nu_0/2, \nu_0 h_0^2/2\right.\right)\; d(a_\alpha, h_\alpha^2), \\ &= t_{\nu_0}\left(\alpha \left| a_0, \left(\frac{\kappa_0+1}{\kappa_0}\right)\nu_0 h_0^2\right.\right), \end{aligned} \tag{13}$$

where the first expectation is taken with respect to the unknown hyperprior, G_α, which equals $G_{\alpha,0}$.

Other than the efficient market hypothesis argument that the average level of skill over the population is zero ($a_0 = 0$), there is little theory to guide us as to what the population distribution for the alphas should look like. Instead, we intentionally set the degrees of freedom, ν_0, to be less than two in order for the variance of the prior predictive distribution of the alphas to be undefined. With an undefined variance our initial guess at the distribution of skill is very diffuse and uninformative.[9]

For the prior of the variances we let $G_{\sigma,0} \equiv \text{Gamma}\left(h_\sigma^2 \middle|\; c_0, 1/b_0\right)$, with shape, $c_0 = 0.3$, and scale, $b_0 = 0.001$. We also set the scale of the conditional, inverse-gamma, prior in Eq. (6) equal to $\nu_\sigma = 0.5$. Dubey (1970) shows that a gamma hyperprior to the conditional inverse-gamma distribution results in the proper beta prime, prior predictive, distribution

$$E_{G_{\sigma,0}}[E_{G_\sigma}[\pi_\sigma(\sigma^2|G_\sigma)]] = \frac{\left(\frac{\sigma^2}{w}+1\right)^{c_0+\nu_\sigma/2}\left(\frac{\sigma^2}{w}\right)^{c_0-1}}{w\;\text{Beta}(c_0, \nu_\sigma/2)}, \tag{14}$$

where $w = \nu_\sigma/(2b_0)$ and $\text{Beta}(c_0, \nu_\sigma/2)$ is the beta function.

For the chosen values of c_0, b_0 and ν_σ, the expected value of the proper beta prime is undefined. However, its median is 402.99 and its lower and upper quantile are 26.5 and 7966.6, respectively. Hence, our initial guess at how the variances are distributed over the funds and their regimes is very diffuse. This ensures that the posterior predictive distribution of the variances will be based on the longitudinal return data.

In the hierarchical prior for the q_is, Eq. (8) is equivalent to the distribution of a j_ith order statistic for a random sample of k_i uniform draws. By the adding-up property of the Bernstein polynomial, the prior predictive for the q_is is thus a uniform distribution on the unit interval

$$E_{G_{q,0}}[E_{G_q}[\pi_q(q|G_q)]] = \text{Uniform}(q|0,1), \tag{15}$$

[9]For our nonparametric hierarchical prior to have a proper prior predictive distribution the DP's base distribution must be a proper distribution. As a result we cannot use a Jeffereys prior for the base distribution.

regardless of the value of k_i. The base distribution, $G_{q,0}$, is characterized by

$$j_i|k_i \sim \text{Uniform}(1,\ldots,k_i), \tag{16}$$

$$k_i - 1 \sim \text{Geometric}(\xi_0). \tag{17}$$

Even though the value of k_i does not affect the prior predictive uniform distribution, its does affect the variation in the random realizations from the nonparametric hierarchical prior. Small values of k_i (a wider beta kernel) restrict the variation in the random realizations, whereas a large k_i (a tighter beta kernel) enhances the variation.

Shrinkage is also controlled by the value of k_i. When $k_i = 1$ there is no local shrinkage in a cluster's q_is, whereas, when $k_i \to \infty$, there is complete local shrinkage. Given these properties we desire more weight be placed on larger values of k_i, so we set $\xi_0 = 1/200$ such that $Pr(k_i = 1) = 1/200$ and $E[k_i] = 200$.

Besides using these specific base distributions for the nonparametric hierarchical priors in our empirical analysis we also use them as the known hyperpriors of the parametric hierarchical priors. The prior predictives are then the same for both the parametric and nonparametric hierarchical priors.

3.2 Posterior DP

After the returns from the panel of mutual funds are observed the posterior of the hierarchical priors are calculated by updating the base distributions of the hyperprior's DP. Since we are interested in the behavior of skill it would be natural to discuss the posterior distribution of the hyperparameters $a_{\alpha,im}$ and $h^2_{\alpha,im}$. However, their posterior has more moving parts and is more complicated than the description of $h^2_{\sigma,im}$ posterior so we only explain the posterior in terms of the variance's hyperparameters (see the online Appendix for complete details on all the posteriors).

Suppose it were possible to observe the hyperparameters of all J funds and their $\mathcal{N}_{\mathcal{A}} = \sum_{i=1}^{J} s_{i,T_i}$ total in-sample regimes.[10] In terms of the hyperparameters for the prior of the variances let $\mathcal{H}_{\sigma,S_T} = \{\boldsymbol{h}_{\sigma,is_{T_i}}\}_{i=1}^{J}$, where $\boldsymbol{h}_{\sigma,is_{T_i}} = (h^2_{\sigma,i1},\ldots,h^2_{\sigma,is_{iT_i}})'$ contains the $\mathcal{N}_{\mathcal{A}}$ scale hyperparameters from all the funds and their in-sample regimes. If we could actually observe $\mathcal{H}_{\sigma,S_T}$ it would contain $\mathcal{N}_{\mathcal{A}}$ independent realizations from the unknown hyperprior, G_σ.

[10] In practice the hyperparameters are never observed, hence, their uncertainty will need to be integrated away after the information found in the return data of the funds and their regimes is used to quantify their posterior distribution.

Combining the hypothetical hyperparameter data in $\mathcal{H}_{\sigma,S_T}$ with the conjugate DP prior for G_σ results in the posterior DP hyperprior

$$G_\sigma \,|\mathcal{H}_{\sigma,S_T} \sim DP\left(\eta_\sigma + \mathcal{N}_\mathcal{A}, \frac{\eta_\sigma G_{\sigma,0} + \sum_{i=1}^{J}\sum_{m=1}^{s_{i,T_i}} \delta_{h^2_{\sigma,im}}}{\eta_\sigma + \mathcal{N}_\mathcal{A}}\right), \tag{18}$$

where $\delta_{h^2_{\sigma,im}}$ is a degenerative distribution at the location $h^2_{\sigma,im}$. Since the concentration parameter of the posterior DP equals $\eta_\sigma + \mathcal{N}_\mathcal{A}$, the uncertainty around G_σ has declined. The base distribution of the posterior DP distribution equals

$$\frac{\eta_\sigma G_{\sigma,0} + \sum_{i=1}^{J}\sum_{m=1}^{s_{i,T_i}} \delta_{h^2_{\sigma,im}}}{\eta_\sigma + \mathcal{N}_\mathcal{A}},$$

which illustrates how the initial guess, $G_{\sigma,0}$, has been updated with the empirical distribution of the "observed" hyperparameters, $M_{S_T}^{-1}\sum_{i=1}^{J}\sum_{m=1}^{s_{i,T_i}} \delta_{h^2_{\sigma,im}}$.

From the base distribution of the posterior DP we can see that as long as η_σ is finite every occurrence of the hyperparameters, $h^2_{\sigma,im}$, adds more empirical information to the guess of G_σ, while giving less weight to the initial guess, $G_{\sigma,0}$. If the DP prior's concentration parameter were infinite ($u_\bullet = 1$), the variance of the DP would be zero and one would learn nothing from the observed data.

At the other end of the spectrum is $\eta_\bullet = 0$ ($u_\bullet = 0$). In this situation every realization of the hyperparameter has the exact same value equal to a single draw from $G_{\sigma,0}$. When this is the case the nonparametric hierarchical prior is equivalent to the parametric hierarchical prior having the hyperprior $G_{\sigma,0}$.

3.3 Clustering

According to the base distribution of the posterior DP in Eq. (18), a realization from the posterior is a draw from either the initial base distribution, $G_{\sigma,0}$, or from the discrete collection of $\mathcal{N}_\mathcal{A}$ degenerative distributions, $\delta_{h^2_{\sigma,im}}$. As explained in Section 3.2 drawing from the discrete distribution depends on the value of η_σ. These draws create ties between the realizations, and, hence, creates $\mathcal{K}_\sigma$ clusters consisting of the unique hyperparameters, $h^{2*}_{\sigma,c}$, $c = 1, \ldots, \mathcal{K}_\sigma$, where $\mathcal{K}_\sigma \leq \mathcal{N}_\mathcal{A}$.[11]

The posterior DP in Eq. (18) can be rewritten in terms of the clusters as

$$G_\sigma \,|H^*_\sigma, \boldsymbol{n}_\sigma \sim DP\left(\eta_\sigma + \mathcal{N}_\mathcal{A}, \frac{\eta_\sigma G_{\sigma,0} + \sum_{c=1}^{\mathcal{K}_\sigma} n_{\sigma,c}\delta_{h^{2*}_{\sigma,c}}}{\eta_\sigma + \mathcal{N}_\mathcal{A}}\right), \tag{19}$$

[11] This clustering is similar to the model-based clustering of Frühwirth-Schnatter & Kaufmann (2008) except the clustering in our DP is in the hyperparameters and the number of clusters, $\mathcal{K}_\sigma$, is unknown.

where $H^*_\sigma = (h^{2*}_{\sigma,1}, \ldots, h^{2*}_{\sigma,\mathcal{K}_\sigma})'$ is a vector comprised of the $\mathcal{K}_\sigma$-uniquely valued $h^2_{\sigma,im}$s, and $\boldsymbol{n}_\sigma = (n_{\sigma,1}, \ldots, n_{\sigma,\mathcal{K}_\sigma})'$, where $n_{\sigma,c}$ is the number of $h^2_{\sigma,im}$s equal to $h^{2*}_{\sigma,c}$. It follows that $\sum_{c=1}^{\mathcal{K}_\sigma} n_{\sigma,c} = \mathcal{N}_\mathcal{A}$.

If the membership to the cth cluster includes a large number of fund-regimes then $n_{\sigma,c}$ will be large relative to $\mathcal{N}_\mathcal{A}$. Then according to Eq. (19), the cth cluster will have a greater chance of a new fund-regime being assigned to it. A nice manageable number of groups occurs, and a sparse, parsimonious, mixture representation of the nonparametric prior is found in

$$\begin{aligned} E_{G_{\sigma,0}}[E_{G_\sigma}[\pi_\sigma(\sigma^2|G_\sigma, H^*_\sigma, \boldsymbol{n}_\sigma)]] &= \frac{\eta_\sigma}{\eta_\sigma + \mathcal{N}_\mathcal{A}} \int IG(\sigma^2|\nu_\sigma/2, \nu_\sigma h^{2*}_\sigma/2)\, dG_{\sigma,0}\left(h^{2*}_\sigma\right) \\ &\quad + \sum_{c=1}^{\mathcal{K}_\sigma} \frac{n_{\sigma,c}}{\eta_\sigma + \mathcal{N}_\mathcal{A}} IG(\sigma^2|\nu_\sigma/2, \nu_\sigma h^{2*}_{\sigma,c}/2). \end{aligned} \tag{20}$$

By partitioning the funds and their regimes into large, small, and possibly singleton ($n_{\sigma,c} = 1$) groups where the members have similarly distributed fund-regime parameters, the DP hyperprior is robust to global shrinkage. Highly skilled funds and extraordinary performing trading strategies are able to speak for themselves in determining their group. Members then share in the information found in their group and the regime parameters shrink locally towards the group's hyperparameters.[12]

Going forward we drop the expectation operators from the predictive distributions knowing that the prior distributions are unknown and estimated with the expected value of their posterior. For instance, we denote the posterior predictive distribution of the variances as

$$\pi_\sigma(\sigma^2|\mathcal{Y}) = \int E_{G_{\sigma,0}}[E_{G_\sigma}[\pi_\sigma(\sigma^2|G_\sigma, \eta_\sigma, H^*_\sigma, \boldsymbol{n}_\sigma)]]\pi(\eta_\sigma, H^*_\sigma, \boldsymbol{n}_\sigma|\mathcal{Y})\, d\eta_\sigma dH^*_\sigma d\boldsymbol{n}_\sigma.$$

4 Posterior simulation

In this section we describe our approach to sampling from the posterior distribution. We divide the unknown change point processes, the regime parameters, and the unknown prior distributions into natural blocks and sample each block from its conditional posterior distribution. For a thorough and complete description of the sampler and the posterior distributions please refer to the online Appendix.

[12]The sampler in Maheu & Song (2014) requires the hierarchical prior to be the joint conjugate prior $\pi(\mu, \sigma^2|a_\mu, h^2_\mu, h^2_\sigma)$, where $(a_\mu, h^2_\mu, h^2_\sigma) \sim G_{\mu,\sigma}$. Because of the Bayesian preference for sparsity such multidimensional DP priors for the hyperprior may have fewer posterior clusters, and hence, possibly be less flexible than our priors.

There are two types of blocks: within-fund blocks and across-fund blocks. In the joint likelihood of Eq. (3) we see the within-fund blocks are conditionally independent across funds and consequently can be parallelized. For a given fund, these blocks include drawing the change point process, the transition probability, and the fund-regime parameters. The across-fund parameters are associated with the nonparametric hierarchical priors. They involve drawing the hyperparameters and concentration parameters. These across-fund blocks are conditionally independent across hyperparameters.

4.1 Sampling the within-fund unknowns

Draws of the in-sample, change point process, S_{i,T_i}, from the conditional posterior distribution, $\pi(S_{i,T_i}|Y_{iT_i}, \boldsymbol{\alpha}_i, \boldsymbol{\sigma}_i^2, q_i)$, are made with the forward-filter, backward-smoother, sampler of Chib (1996). As we have already pointed out the draws of S_{i,T_i} are parallelized over the J funds. Note that these latent change point draws only go to the end of the return series, Y_{i,T_i}. To sample the out-of-sample change points we apply the same forward-backward approach up to the M_ith regime.

Given the draw of S_{i,T_i}, and the hyperparameters, j_i and k_i, we then draw q_i from

$$\pi(q_i|\mathcal{T}_i, s_{i,T_i}, j_i, k_i) = \text{Beta}\left(q_i \left| j_i + s_{i,T_i} - 1, k_i - j_i + 1 + (\mathcal{T}_i - s_{i,T_i})\right.\right), \tag{21}$$

where $s_{i,T_i} - 1$ is the number of change-points out of $\mathcal{T}_i - 1$ independent Binomial trails (the minus one accounts for $s_{i,\tau_i} = 1$ with probability one).

To draw the elements of $\boldsymbol{\alpha}_i$ and $\boldsymbol{\sigma}_i^2$ we condition on the current draw of S_{i,T_i} and the hyperparameters, $(a_{\alpha,im}, h_{\alpha,im}^2)$ and $h_{\sigma,im}^2$, $m = 1, \ldots, M_i$. Given the conjugate hierarchical prior for the variances found in Eq. (6), σ_{im}^2 is drawn from the conditional posterior by sampling from the marginal distribution

$$\pi\left(\sigma_{im}^2 \left| \{y_{i,t} : s_{i,t} = m\}, \alpha_{im}, h_{\sigma,im}^2\right.\right) \propto \prod_{t:s_{i,t}=m} N\left(y_{i,t} \left| \alpha_{im}, \sigma_{im}^2\right.\right) \text{IG}\left(\sigma_{im}^2 \left| \nu_\sigma/2, \nu_\sigma h_{\sigma,im}^2\right.\right).$$

The α_{im}s are then drawn from the conditional posteriors

$$\pi(\alpha_{im}|\{y_{i,t} : s_{i,t} = m\}, \sigma_{im}^2, a_{\alpha,im}, h_{\alpha,im}^2) \propto \prod_{t:s_{i,t}=m} N(y_{i,t}|\alpha_{im}, \sigma_{im}^2) N(\alpha_{im}|a_{\alpha,im}, h_{\alpha,im}^2).$$

Note that when the ith fund's regime, m, is greater than s_{i,T_i} the posteriors for α_{im} and σ_{im}^2 have no return data to condition on. To draw the regime parameters for these $M_i - s_{i,T_i}$ out-of-sample regimes we sample from the posterior conditional priors

$$\sigma_{im}^2 \sim \text{IG}(\sigma_{im}^2|\nu_\sigma/2, \nu_\sigma h_{\sigma,im}^2), \tag{22}$$

$$\alpha_{im} \sim N(y_{i,t}|\alpha_{im}, \sigma_{im}^2)\, N(\alpha_{im}|a_{\alpha,im}, h_{\alpha,im}^2), \tag{23}$$

where the hyperparameters, $h^2_{\sigma,im}$, $a_{\alpha,im}$, and $h^2_{\alpha,im}$, for $m > s_{i,T_i}$, are draws from the out-of-sample posterior hyperprior described in Section 4.2.2.

4.2 Sampling the across-fund unknowns

Given each fund's posterior draw of $\boldsymbol{\alpha}_i$, $\boldsymbol{\sigma}^2_i$, and q_i, the sampler moves on to drawing the across-fund unknowns associated with the nonparametric hierarchical representation of π_α, π_σ, and π_q. All three priors have hierarchical conjugate hyperpriors for the DP priors and their base distributions. This allows the hyperparameters to be quickly and efficiently drawn with the two-step DP sampler of West et al. (1994), MacEachern & Müller (1998), and described by Neal (2000) in his Algorithm 2.

The DP two-step sampler first sequentially assigns every regime of every fund a random hyperparameter and in the process partitions the fund-regimes into groups having the same hyperparameters. In the second step the hyperparameters of each regime group are independently drawn.

4.2.1 Transition probability hyperparameters

For the prior of the transition probabilities, the first step of the DP sampler sequentially draws the j_i and k_i along with the unknown cluster assignment variable, $z_{q,i}$, for $i = 1, \ldots, J$. The assignment variable, $z_{q,i} = c$, where $c = 1, \ldots, \mathcal{K}_q$, when $j_i = j^*_c$ and $k_i = k^*_c$, and j^*_c, and k^*_c are the unique hyperparameters of the cth cluster. These draws and assignments are all done by applying the Chinese restaurant process sampling algorithm to the j_is and k_is (see Teh (2010) for details on the Chinese restaurant process).

In the second step, separate draws are made from the posterior $\pi(j^*_c, k^*_c | \{q_i : z_{q,i} = c\})$, for $c = 1, \ldots, \mathcal{K}_q$. Since this conditional posterior distribution is not an analytical distribution, we design an efficient Metropolis-Hasting (MH) sampling scheme to draw from this distribution. The technical details of this MH sampler are described in the online Appendix.

4.2.2 Regime hyperparameters

To sample the hyperparameters $a_{\alpha,im}$, $h^2_{\alpha,im}$ and $h^2_{\sigma,im}$, we partition the hyperparameters, the alphas, α_{im}, and the variances, σ^2_{im}, into in-sample ($m \leq s_{i,T_i}$) and out-of-sample ($s_{i,T_i} < m \leq M_i$) regimes. Because M_i exceeds s_{i,T_i}, our sampler always has out-of-sample unknowns where there is no return data.[13] As a result the out-of-sample hyperparameter

[13] If at any point in the sampler the inequality, $s_{i,T_i} < M_i$, did not hold we increased M_i and started the entire sampler over.

draws are realizations from the hyperprior's posterior predictive distribution.

Let the in-sample hyperparameters, $a_{\alpha,im}$, $h^2_{\alpha,im}$, and, $h^2_{\sigma,im}$, where $m = 1, \ldots, s_{i,T_i}$, and $i = 1, \ldots, J$, be stored in $\mathcal{A}_{\alpha,s_T}$, $\mathcal{H}_{\alpha,s_T}$, and $\mathcal{H}_{\sigma,s_T}$, respectively. Each set contains the $\mathcal{N}_{\mathcal{A}}$ in-sample hyperparameters. Also let the $\mathcal{N}_{\mathcal{A}}$-element set of in-sample alphas and variances be defined as $\boldsymbol{A}_{s_T} = \{\alpha_{im} : m = 1, \ldots, s_{i,T_i}, \quad i = 1, \ldots, J\}$ and $\boldsymbol{\Sigma}_{s_T} = \{\sigma^2_{im} : m = 1, \ldots, s_{i,T_i}, \quad i = 1, \ldots, J\}$, respectively.

The two base distributions

$$G_{\alpha,0} \equiv \text{NIG}(a_\alpha, h^2_\alpha | a_0, h^2_\alpha/\kappa_0, \nu_0/2, \nu_0 h^2_0/2), \qquad G_{\sigma,0} \equiv \text{Gamma}(h^2_\sigma | c_0, 1/b_0),$$

are conjugate to the normally distributed conditional prior in Eq. (4) and the inverse-gamma distributed conditional prior in Eq. (6), respectively. Hence, drawing the in-sample hyperparameters from $\pi(\mathcal{A}_{\alpha,s_T}, \mathcal{H}_{\alpha,s_T} | \boldsymbol{A}_{s_T}, \eta_\alpha)$ and $\pi(\mathcal{H}_{\sigma,s_T} | \boldsymbol{\Sigma}_{s_T}, \eta_\sigma)$ is a straight-forward application of the two-step DP sampler (see the Internet Appendix for the technical details of these two-step draws).

Since the out-of-sample hyperparameters are exchangeable the order in which they are drawn does not matter. So for a tractable description of the out-of-sample draws one can think of the sampler starting with the first fund and then sequentially sampling the first fund's out-of-sample hyperparameters for the regimes, $s_{1,T_1} + 1, \ldots, M_1$, before moving on to drawing the out-of-sample hyperparameters for the second fund. The sampler then continues sequentially drawing the out-of-sample hyperparameters, fund by fund, until the hyperparameters have been drawn for all J mutual funds.

Formally, the draw of the ith fund's mth out-of-sample hyperparameters, $(a_{\alpha,im}, h^2_{\alpha,im})$, where $m > s_{i,T_i}$, are sampled from the conditional posterior predictive distribution

$$\frac{1}{\eta_\alpha + M_{s_T} + M_{im}} \Big[\eta_\alpha NIG\left(a_{\alpha,im}, h^2_{\alpha,im} \,\middle|\, a_0, h^2_\alpha/\kappa_0, \nu_0/2, \nu_0 h^2_0/2\right) + \sum_{i'<i} \sum_{m'=1}^{M_{i'}} \delta_{(a_{\alpha,i'm'}, h^2_{\alpha,i'm'})} + \sum_{l=1}^{m-1} \delta_{(a_{\alpha,il}, h^2_{\alpha,il})} \Big], \tag{24}$$

where M_{im} is the number of out-of-sample hyperparameters drawn before sampling $(a_{\alpha,im}, h^2_{\alpha,im})$.[14] Posterior draws of the out-of-sample hyperparameters, $h^2_{\sigma,im}$, where $m > s_{i,T_i}$, are similarly made fund by fund (see the online Appendix for the exact form of the posterior distribution).

[14] Note that the out-of-sample hyperparameters will cluster into groups having the same unique values, $(a^*_{\alpha,c}, h^{2*}_{\alpha,c})$, $c = 1, \ldots, \mathcal{K}_\alpha$. Therefore, we could have written Eq. (24) in terms of the unique hyperparameter values, however, this would have required additional notation and a more involved explanation.

4.3 Concentration parameter draws

Posterior draws of the concentration parameters, η_μ, η_σ and η_q, are all made using the same MH algorithm. The density for the MH proposal draw $\eta'_\bullet$ is $\eta'_\bullet = \eta_\bullet u_\bullet/(1-u_\bullet)$ where $u_\bullet \sim \text{Uniform}(0,1)$. Given the Log-Logistic prior for the concentration parameter it follows that $\pi(\eta'_\bullet|\eta_\bullet) = \eta_\bullet/(\eta_\bullet + \eta'_\bullet)^2$. Step-by-step details of the MH sampler are found in the online Appendix.

4.4 Posterior predictive

To compute the posterior predictive distributions we integrate away the uncertainty found in the hyperparameters of the hierarchical prior and the DP's concentration parameters by numerically averaging the conditional distribution of the parameter over the posterior draws of these unknowns. For example, the predictive distribution for the variances is calculated as

$$\begin{aligned}\pi_\sigma(\sigma^2|\mathcal{Y}) \;\approx\; R^{-1}\sum_{r=1}^{R}\Bigg[&\frac{\eta_\sigma^{(r)}}{\eta_\sigma^{(r)}+M_{S_T}^{(r)}} f_{PBP}(\sigma^2|c_0,\nu_\sigma/2,1,w)\\ &+\sum_{c=1}^{\mathcal{K}_\sigma^{(r)}}\frac{n_{\sigma,c}^{(r)}}{\eta_\sigma^{(r)}+M_{S_T}^{(r)}} f_{IG}\left(\sigma^2\,\middle|\,\nu_\sigma/2,\nu_\sigma h_{\sigma,c}^{2*(r)}/2\right)\Bigg],\end{aligned} \tag{25}$$

which averages the conditional posterior predictive distribution from Eq. (20) over the R posterior draws $h_{\sigma,c}^{2*(r)}$, $\eta_\sigma^{(r)}$ and $M_{S_T}^{(r)}$, $r = 1,\ldots,R$. Similar calculations are performed for the posterior predictive densities, $\pi_\alpha(\alpha|\mathcal{Y})$, and $\pi_q(q|\mathcal{Y})$ (see the online Appendix for these formulas).

5 Mutual fund application

An important question mutual fund investors ask is, do above market returns by a mutual fund today lead to above market returns in the future? Along with an answer to this question investors would also like to know if the performance of a fund is linked to the success of its manager. These and other performance related questions have been addressed in the academic and profession literature dating back to Jensen (1968).[15] Bollen & Whaley (2009) apply a single-change-point model to a cross-section of hedge funds, but, to our knowledge a multiple-change-point panel model has never been used to measure mutual fund skill and its persistence.

[15] See Elton & Gruber (2013) for a review of the mutual fund performance literature.

To measure skill and its persistence we analyze mutual fund performance with the multiple-change-point, four-factor-risk, model

$$y_{i,t} = \alpha_{i,s_{i,t}} + \beta_{R,i,s_{i,t}}\text{RMRF}_t + \beta_{S,i,s_{i,t}}\text{SMB}_t + \beta_{H,i,s_{i,t}}\text{HML}_t + \beta_{M,i,s_{i,t}}\text{MOM}_t + \sigma_{i,s_{i,t}}\epsilon_{i,t}, \quad (26)$$

where $y_{i,t}$ is the risk-free adjusted, gross rate of return generated by the ith mutual fund in month t. The risk factors are Fama & French's (1993) three factors of excess market returns, RMRF_t, market size, SMB_t, and book-to-market, HML_t, plus Carhart's (1997) momentum portfolio factor, MOM_t. The intercept, $\alpha_{i,s_{i,t}}$, is a change point version of the Jensen (1968) alpha that allows a fund's level of performance to change with its change point process, $s_{i,t}$.

We assume the prior for $\alpha_{i,s_{it}}$ and the risk factor parameters, $\beta_{R,i,s_{it}}$, $\beta_{S,i,s_{it}}$, $\beta_{H,i,s_{it}}$, and $\beta_{M,i,s_{it}}$, consists of the independent marginals denoted by π_α, π_{β_R}, π_{β_S}, π_{β_H}, and π_{β_M}, respectively. Each prior is unknown and modeled with the nonparametric hierarchical prior defined in Eq. (4), (5), and (10). In this paper we only discuss our findings for the distributions of the alphas but in future work we plan to report and analyze the distributions of the betas.

5.1 Longitudinal mutual fund data

Our longitudinal data set consists of the monthly mutual fund returns investigated by Jones & Shanken (2005).[16] For ease of comparison we annualized the returns by multiplying the monthly return by twelve so the alphas are yearly above market rates of return. The data is comprised of the returns from every US domestic, equity, mutual fund that existed between January 1961 to June 2001.[17] Each fund has at least a years worth of return data and on average the funds have 77.3 months of returns. Survivorship bias is not a issue since the panel includes the returns from all 1,293 funds that did not survive to the end of the sample. In total the data consists of a unbalanced panel of 396,820 monthly observations across 5,136 domestic equity funds.

5.2 Posterior inference

Posterior inference is made using both nonparametric and parametric hierarchical priors. As we explained in Section 3.1, neither approach is at an initial disadvantage since we set the base distribution of the DP hyperpriors equal to the parametric hyperpriors. Both approaches have similar samplers but the parametric model only needs to draw a single

[16] We would like to thank Chris Jones for providing us with their data.

[17] Funds were excluded if they made substantial investments in asset classes other than domestic equities.

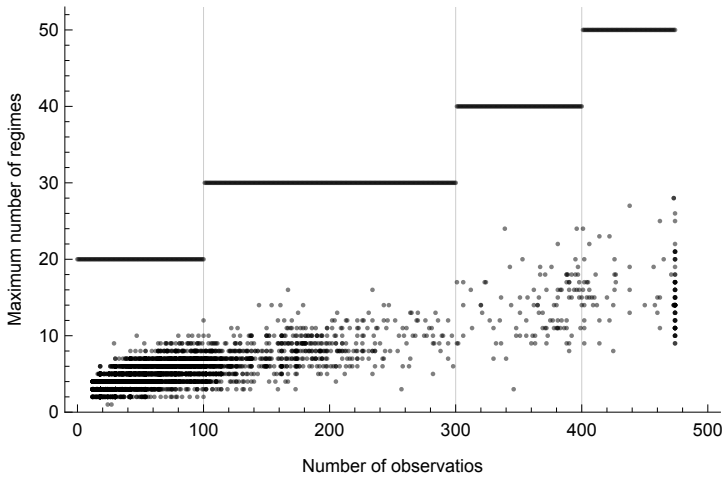


Figure 1: The four horizontal lines are the pre-set thresholds for the maximum number of in- and out-of-sample regimes, M_i, plotted against the length of the fund's history, $\mathcal{T}_i$. Each point is the largest MCMC draw of s_{i,T_i} under the nonparametric priors.

global hyperparameter for all the fund-regimes, whereas the nonparametric model can draw a different hyperparameter value for each fund-regime.

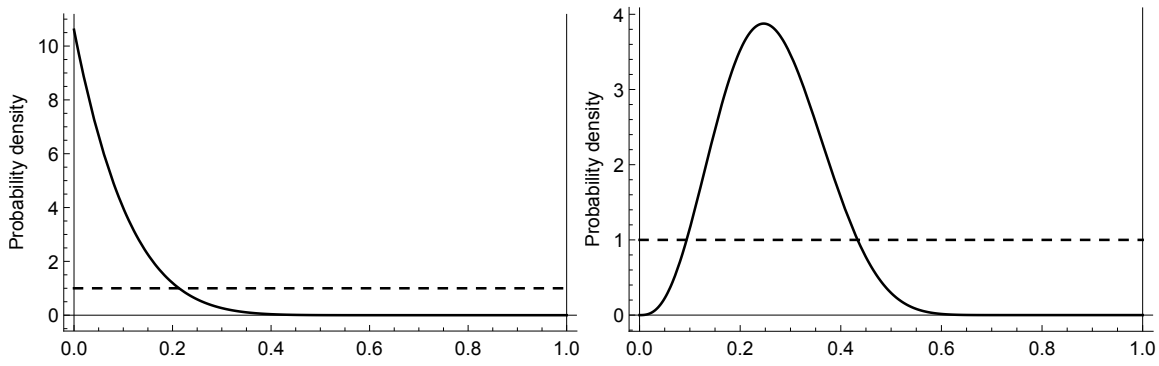
For both models we start collecting the posterior draws after discarding the first 30,000 MCMC draws. We run the sampler for the nonparametric priors first and use its last draw to initialize the unknowns in the parametric hierarchical prior. Noticeable trending was observed in the initial draws of the nonparametric approach but by the end of the burnin period the sampler had settled down and was generating well behaved draws of the unknowns. We then iterate each sampler for 60,000 more sweeps keeping every sixtieth draw to construct a random sample of 1,000 posterior draws.[18]

5.3 Number of regimes

In both the nonparametric and parametric case the maximum number of fund-regimes, M_i, is set so as not to restrict the number of in-sample regimes. We use four different values, 20, 30, 40 and 50 fund-regimes where the funds with the shortest histories have $M_i = 20$ and those with the longest have $M_i = 50$ (this is visually shown in Figure 1). By discriminating on the number of returns we are able to speed up the sampler.

To put to rest any doubts that our posterior results have been affected by the preset value of the M_i, in Figure 1 we plot M_i and the largest draw of s_{i,T_i} under the nonparametric

[18]Because the functional form of the conditional posterior densities have known analytical formulas, 1,000 draws is large enough to accurately represent these posterior densities after marginalizing out the unknowns over the 1,000 draws.



(a) u_q (b) u_α

Figure 2: Posterior distribution of the transformed concentration parameter u_q in panel (a) and u_α in panel (b).

prior against $\mathcal{T}_i$ (the technical online appendix contains the same figure for the parametric prior). In the figures the largest draw of s_{i,T_i} never gets close to M_i. It is interesting to point out that the maximum draw of s_{i,T_i} range from two to almost thirty regimes. This suggests Bollen & Whaley (2009) single-change-point model may be too restrictive.

Under the nonparametric priors there are an average of 12,390 different fund-regimes over all the funds, which produces an average of 2.4 regimes per fund. The average fund-regime then lasts around three years. Under the parametric priors the mean number of total fund-regimes drops to 11,914. This drop is statistically different since the posterior densities of the number of fund-regimes from the two approaches do not overlap.[19] Hence, we conclude that the parametric priors lead to fewer fund-regimes than the nonparametric priors.

5.4 Model comparison

Because the nonparametric hierarchical prior nests the parametric when the concentration parameter is zero, the Savage-Dickey (SD) density ratio of Dickey (1971) can be used to quickly compute the Bayes factor in favor of the parametric prior ($\eta_\bullet = 0$). In terms of the transformed concentration parameter, $u_\bullet = \eta_\bullet/(1+\eta_\bullet)$, the SD density ratio is

$$\text{BF}(u_\bullet = 0) \equiv \frac{\pi(u_\bullet = 0|\mathcal{Y})}{\text{Uniform}(u_\bullet|0,1)},$$

where $\pi(u_\bullet|\mathcal{Y})$ is the posterior distribution of $u_\bullet$.

[19]The posterior densities of the number of fund-regimes can be found in the online Appendix.

In Figure 2 we plot in panel (a) the Rao-Blackwellized posterior distribution of the transformed concentration parameters u_q for the transition probabilities, q_i, and in panel (b), the posterior of u_α. The uniform prior for the $u_\bullet$s are indicated in the plots by the horizontal dashed line at one. Since the prior equals one over its range of zero to one, visually a SD ratio in favor of the parametric prior is the vertical height of the posterior at zero. In panel (a) we see that the odds of the data supporting the parametric hierarchical prior of q_i are slightly better than ten-to-one over the nonparametric prior. On the other hand, in panel (b) the odds in favor of the parametric distribution of alpha is essentially zero.

Panel (b) of Figure 2 shows there is strong empirical evidence in favor of the distribution of the alphas having more than one cluster. The posterior is effectively zero at $u_\alpha = 0$, meaning that a parametric hierarchical prior of the alphas is a poor modeling choice. The density is also effectively zero at $u_\alpha = 1$ so there is essentially zero empirical evidence for each fund-regime having its own unique cluster. Given the near zero probabilities for u_α equaling zero or one, assuming the hyperprior, G_α, is known apriori would be a bad modeling assumption.

5.5 Posterior predictive distributions

In Figures 3, 4, and 5, we plot the nonparametric posterior predictive densities, $\pi_q(q|\mathcal{Y})$, for the transition probabilities, $\pi_\alpha(\alpha|\mathcal{Y})$, for the alphas, and, $\pi_\sigma(\sigma^2|\mathcal{Y})$, for the variances, respectively. Each figure includes the corresponding predictive density under the parametric hierarchical priors. Eq. (25) is used to compute the density in Figure 5 and similar Rao-Blackwell formulas are used to compute the other posterior predictive densities.

The nonparametric predictive distribution of the transition probability plotted in Figure 3 has a posterior median of one cluster; ie., $\mathcal{K}_q = 1$. Given the height of the density in panel (a) of Figure 2 at the orgin, finding there to be only one cluster is not a surprise. Having only one cluster complete sharing of information under both the nonparametric and parametric prior of q occurs.

Both posterior predictive densities of q strongly favor a multiple-change-point process for the parameters of the four-factor risk model. The probability of q being less than 0.006 is essentially zero under both posterior densities of Figure 3. Therefore, skill clearly changes over time and a mutual fund performance model whose alpha is assumed to be constant would amount to a very restrictive prior.

The nonparametric density $\pi_\alpha(\alpha|\mathcal{Y})$ plotted in Figure 4 has a posterior median of four clusters, $\mathcal{K}_\alpha = 4$, with three clear modes at approximately -1.75, 1.5, and 6. Since our

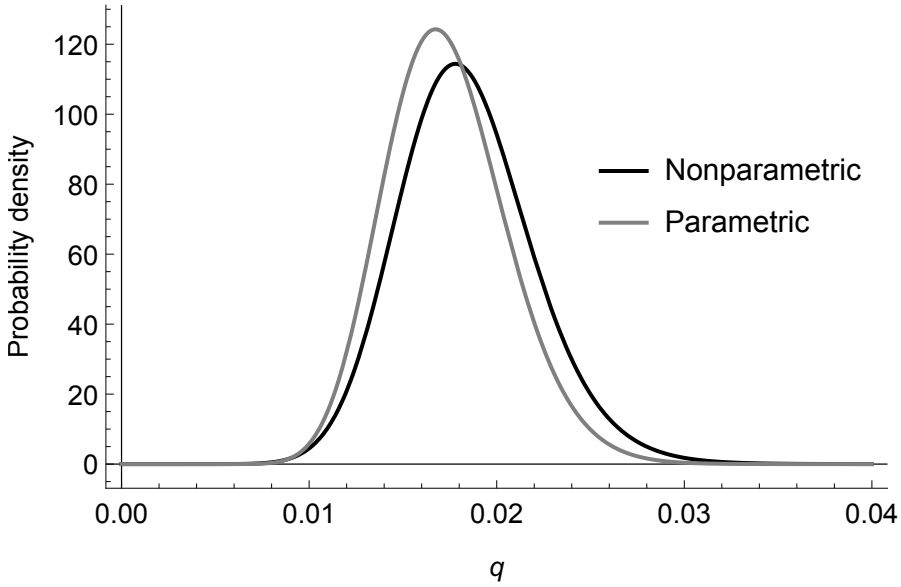


Figure 3: The parametric and nonparametric posterior predictive density of the transition probability, q, $\pi_q(q|\mathcal{Y})$.

goal differs from Malsiner-Walli et al. (2016), and, instead, focuses on flexibly modeling the distribution of the fund-regime alphas, we are unable to make economic statements about these modes.[20] Nonetheless, these modes suggest the average performance levels of three distinct groups, unskilled, break-even, and skilled fund-regimes, respectively.

The break-even mode of 1.5 has economic support in Berk & Green (2004) who find the average actively managed mutual fund charges a fee of 1.5 percent. The secondary modes are also consistent with Berk & Green (2004) and their decreasing return to scale model of fund performance. Under their model an alpha greater than the break-even level leads to investors increasing their investment in the skilled fund. This causes the assets under management of the fund to increase and, because of the decreasing returns to scale, causes its alpha to move to a lower fund-regime. The opposite occurs for an unskilled fund.

Alphas belonging to the same cluster shrink towards the cluster average. In Figure 4 the greatest shrinkage is found in the tight variance around the α_{im}s belonging to the hypothesized break-even cluster. In contrast, the parametric predictive density shows complete sharing and global shrinkage. Alphas from the extraordinary fund-regimes end up being

[20]This is because of identification issues that arise from the label switching problem detailed by Geweke (2007) and Frühwirth-Schnatter (2006).

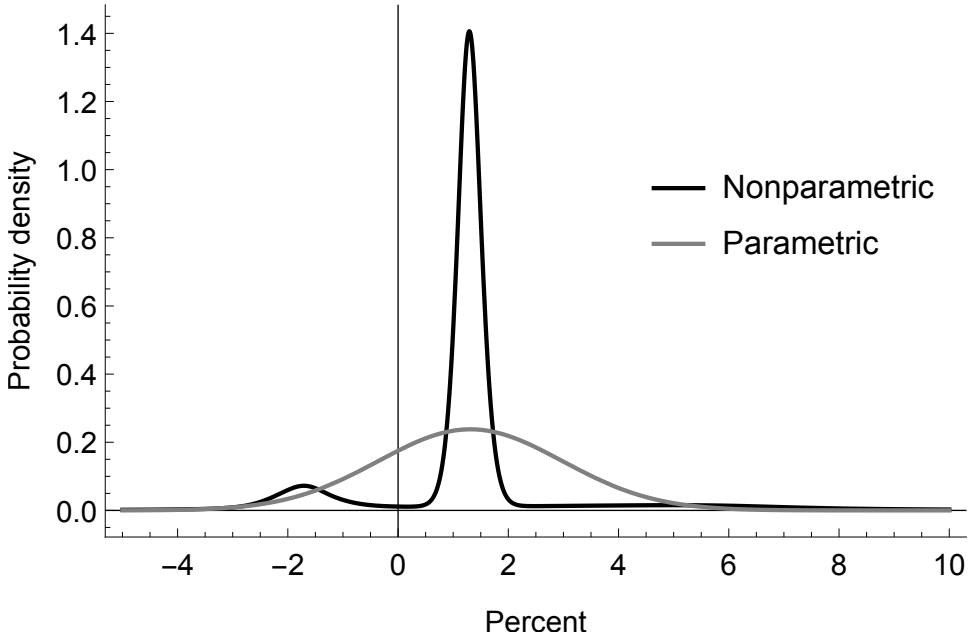


Figure 4: Parametric and nonparametric posterior predictive density of alpha for a future fund-regime, $\pi_\alpha(\alpha|\mathcal{Y})$.

contaminated by the information from the average fund-regime skill level of the entire population and end looking like the break-even regime.

In Figure 5 we plot the nonparametric posterior predictive density of the regime variances. This density has two modes – a primary mode near a variance of one-hundred and a secondary mode near one.[21] Except for the mode near one, the parametric posterior predictive density looks identical to the nonparametric density.

Our working hypothesis for the mode near one is that it belongs to a group of index funds that track the market. By tracking market returns these funds are less risky and have on average a lower variance. In contrast non-market index funds take greater risks in order to outperform the market.

5.6 Posterior fund performance

Each mutual fund in our panel has its own history of posterior smoothed dynamic densities for $\alpha_{i,t}$. Of these we have selected the highest performing fund with the longest tenure to illustrate the smoothed densities of the alphas. The fund is the Fidelity Magellan Fund which opened for business in 1963. Magellan had one of the best known fund manager in

[21] Please refer to the online Appendix for a graph that more clearly shows the mode near zero.

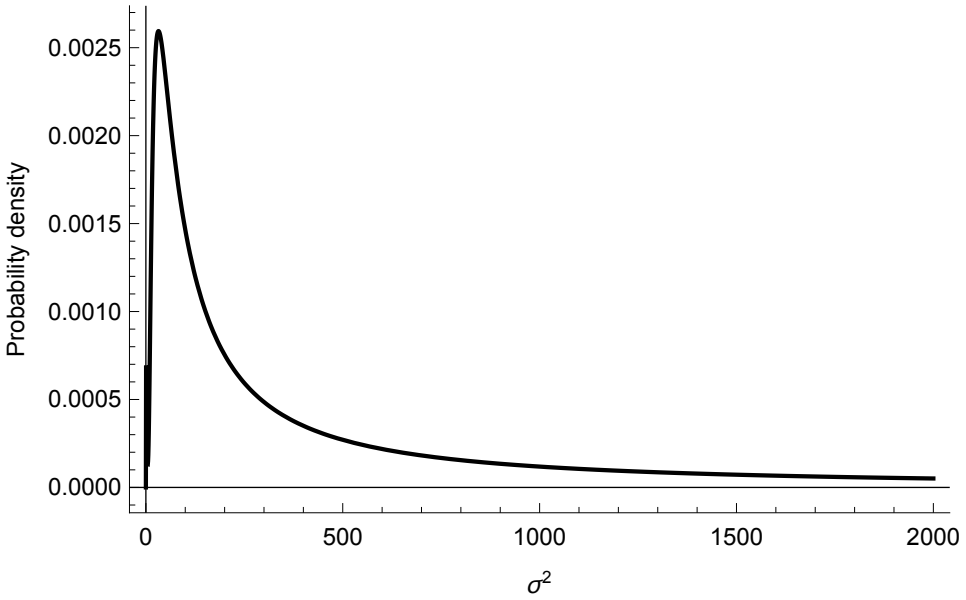


Figure 5: Posterior nonparametric predictive density of a future fund-regimes variance, $\pi_\sigma(\sigma^2|\mathcal{Y})$.

Peter Lynch, who managed the fund for thirteen years from May 1977 to May 1990. Besides Lynch the fund had been managed by five other fund managers over our sample period.

Under the nonparametric priors Magellan experiences on average sixteen regimes over its entire time period, whereas the parametric priors uncovers an average of seventeen regimes. Magellan's probability of switching regimes under both types of hierarchical priors is nearly identical. Under the nonparametric priors the transition probability is 0.022, whereas it is 0.021 for the parametric priors. Both probabilities are only slightly larger than the 0.018 posterior transition probability for the population. Thus, regardless of the hierarchical prior, the duration of a Magellan regime is only slightly longer than the typical fund-regime.

In Figure 6 we plot the posterior median and highest probability intervals (HPD) for the 90, 75, and 50 percentile of Magellan's posterior distribution $\pi_\mu(\alpha_{FM,t}|\mathcal{Y})$ from June 1963 to June 2001 (please refer to the online technical appendix for the calculation of these distributions). In the top panel are the HPDs from the nonparametric priors and in the bottom panel are those for the parametric priors. Each fund manager's tenure is denoted by the five vertical lines with the longest tenure being that of Lynch.

Comparing the top and bottom panels of Figure 6 we see how the global shrinkage caused by the parametric (bottom) approach dampens the exceptional regimes relative to

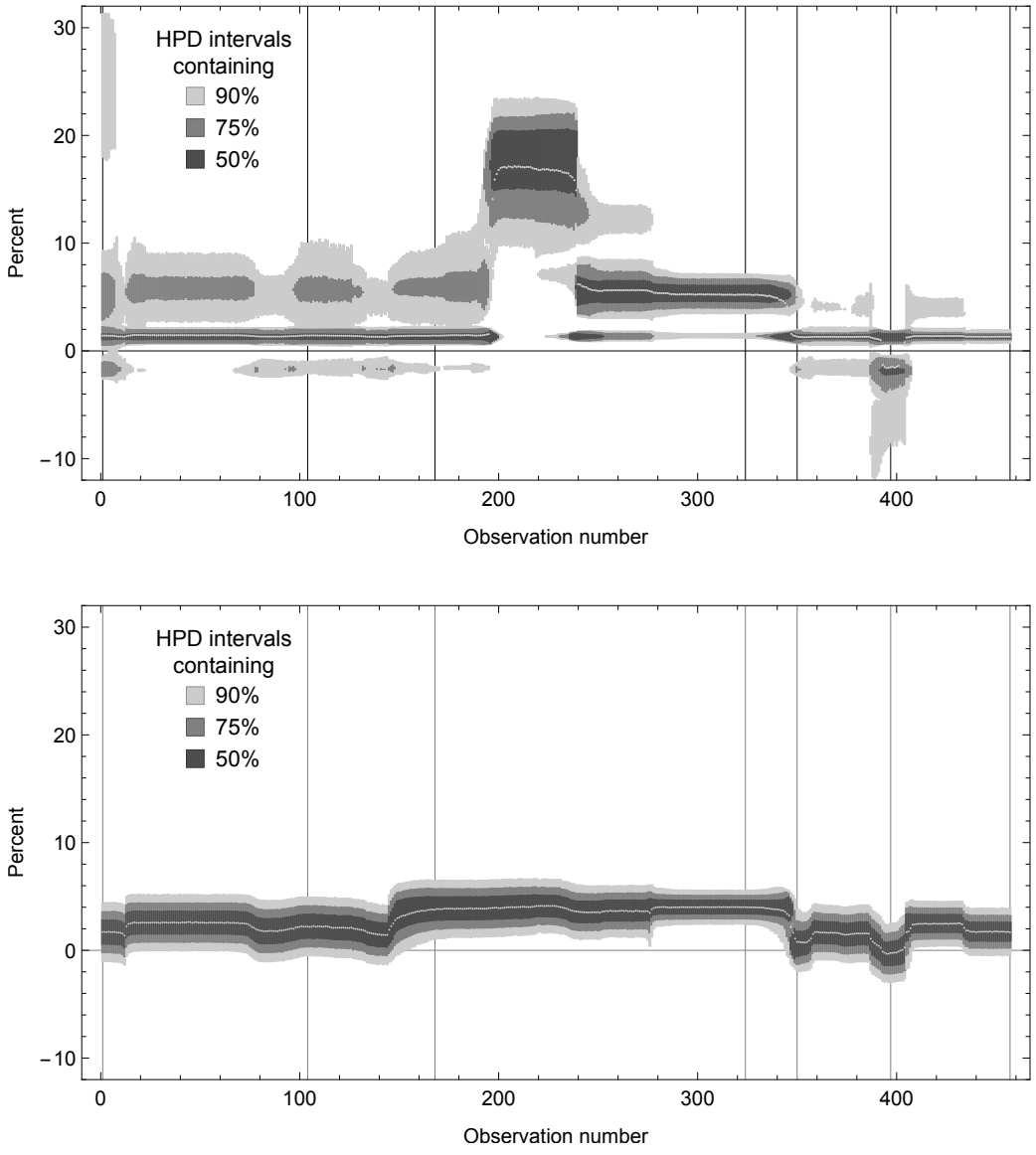


Figure 6: The highest probability density intervals (HPD) and posterior medians (white dots) under the nonparametric prior (top graph) and parametric prior (bottom graph) from the posterior densities $\pi(\alpha_{FM,t}|\mathcal{Y})$, $t = 06/1963, \ldots, 06/2011$ for Fidelity Magellan. The vertical lines at 104 (01/1972), 168 (05/1977), 324 (05/1990), 350 (07/1992), 397 (06/1996), and 457 (06/2001) denote a change in the fund manager. Lynch was the manager from 168 to 324.

the local shrinkage of the nonparametric priors (top). Although the parametric posteriors during the Lynch era (observations 168 to 324) had higher medians and HPD intervals than during the rest of Magellen's history, the intervals shifted up before Lynch. Contrast this with the dramatic increase in the nonparametric posterior's medians and HPDs after Lynch took over. Under the parametric prior any exceptional performance attributable to Lynch is muted and occurred before he became manager.

Except for the instance where the alphas for Magellan were already transitioning to a new regime, changing managers fails to result in any change in fund performance under either the parametric or nonparametric priors. Both panels in Figure 6 suggest Magellan continued to deliver the same alpha before and after a change in the manager.

The nonparametric posterior medians in Figure 6 begin transitioning later and take fewer months to reach their new regime than do the parametric posterior medians. Local shrinkage under the nonparametric approach is the likely reason. Instead of shrinking towards the global average, local shrinkage pulls the medians towards the local average. This explains the quick transition from Lynch's early ordinary performance levels to his extraordinary returns of eighteen to nineteen percent a year.

Figure 7 plots each of the posterior densities $\pi_\mu(\alpha_{FM,t}|\mathcal{Y})$ used to compute the HPDs in the top panel of Figure 6. Magellan's five fund manager's tenures are denoted by different gray scales along with axis ticks. Many of the densities resemble the nonparametric posterior predictive density plotted in Figure 4 with their three modes.[22] This resemblance occurs because the posterior predictive distribution is essentially the prior for Magellan once the prior predictive has been updated with the information contained in the return data of the 5,135 other funds.

There are also densities in Figure 7 that are visibly different from the population distribution of the alphas. These exceptional densities have fewer than three modes and are centered at larger values of alpha. For instance, during the Lynch period the posterior is unimodal at 18%. These episodes are periods of truly exceptional skill and performance. Lynch's investment strategy during this period eliminated the negative and break-even modes. This exceptional performance lasts for nearly three years before the break-even mode reappears and the primary mode retreats to a lower but still exceptional alpha of six percent. This extraordinary shape, relative to the population distribution of skill, is kept for the remainder of Lynch's time as manager and through most of the tenure of the next manager.

[22] Although not shown here this is also the case for Magellan's parametric posterior densities, $\pi_\mu(\alpha_{FM,T_{FM}+1}|\mathcal{Y})$.

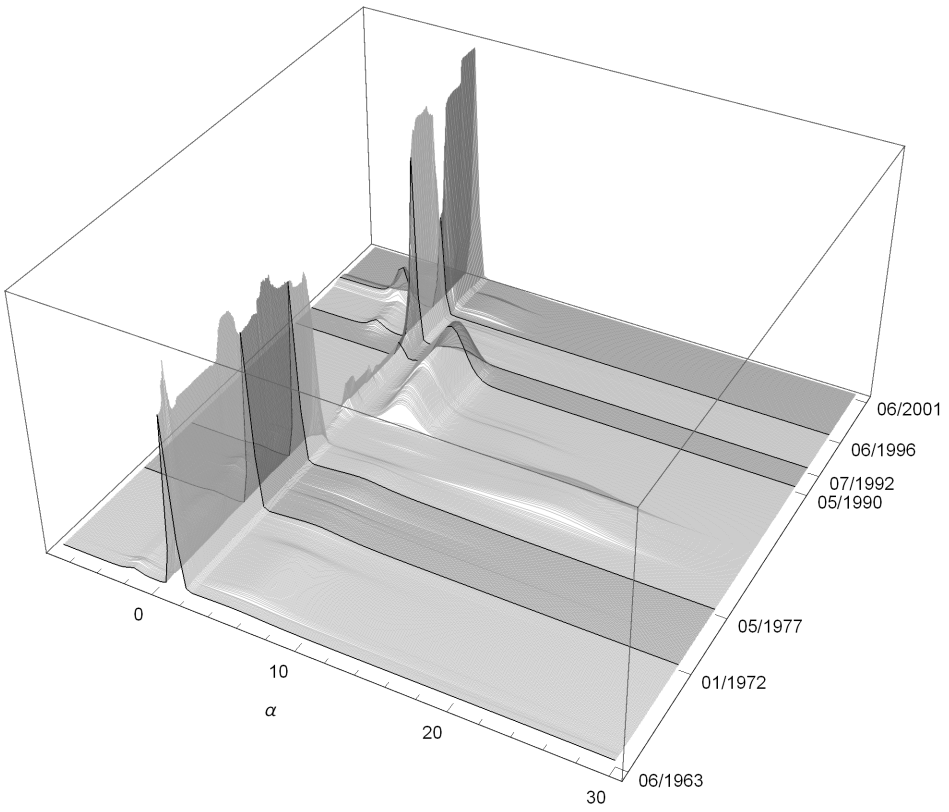


Figure 7: Posterior densities for the alpha of the Fidelity Magellan Fund, $\pi(\alpha_{FM,t}|\mathcal{Y})$, where $t = 06/1963, \ldots, 06/2001$. A change in the shade of gray denotes the end to, and the start of, one of Magellan's six fund managers. These dates are denoted by tics on the x-axis (Lynch's tenure began in 05/1977 and ended in 05/1990).

6 Conclusion

In this paper we extended the change point model to a panel of multiple-change-point processes where the parameters are modeled with nonparametric hierarchical priors. Under nonparametric hierarchical priors our multiple-change-point panel model shares information from across the regimes and panel individuals through the hierarchical prior's hyperparameters and their unknown hyperprior distribution to produce more robust parameter estimates. Our nonparametric hierarchical priors cluster together the individuals and their regimes into groups that have similar behavior and locally shrink the unknown parameters towards the average of the group. This partial sharing and local shrinkage allows extreme regimes and individuals to be identified.

We apply our multiple-change-point panel model and nonparametric hierarchical priors to a longitudinal data set of mutual fund returns to investigate the skill level of mutual funds and how likely they are to maintain this skill into the future. There is overwhelming empirical evidence supporting change-point behavior in the fund-regime parameters. Hence, assuming that mutual fund skill is constant over the history of a fund is a very bad assumption.

Under the nonparametric hierarchical priors we find on average four clusters and three modes for the population distribution of skill; a primary mode at 1.5 percent a year, another mode at -1.75, and a very diffuse mode at around 6 percent. Compared to the parametric prior's single mode at 1.5 percent there is strong empirical evidence against using the parametric hierarchical prior to estimate skill. Having three modes keeps the estimate of skilled and unskilled fund's performance from being masked by shrinking to the average performance level of 1.5 percent.

Our findings for the population carries over to the performance of one of the longest lived skilled funds in the panel, Fidelity Magellan. We find Magellan to have extended periods of time where its performance is average, highly skilled, and also low skilled. Under the parametric prior these periods of extraordinary performance were not that extraordinary. Periods of exceptional skill were hidden by the parametric prior's restrictive single mode assumption for the population distribution of skill.

In future research we aim to design a method capable of finding the funds and their regimes that have similar posterior densities and use them to investigate the structural nature of skill. These and many other interesting questions can be addressed using our nonparametric hierarchical prior with the multiple-change-point panel model.

References

Bassetti, F., Casarin, R. & Leisen, F. (2014), 'Beta-product dependent Pitman–Yor processes for Bayesian inference', *Journal of Econometrics* **180**(1), 49 – 72.

Bauwens, L., Koop, G., Korobilis, D. & Rombouts, J. V. (2015), 'The contribution of structural break models to forecasting macroeconomic series', *Journal of Applied Econometrics* **30**(4), 596–620.

Berk, J. B. & Green, R. C. (2004), 'Mutual fund flows and performance in rational markets', *Journal of Political Economy* **112**, 1269–1295.

Billio, M., Casarin, R., Ravazzolo, F. & Van Dijk, H. K. (2016), 'Interconnections between Eurozone and US booms and busts using a Bayesian panel Markov-switching VAR model', *Journal of Applied Econometrics* **31**(7), 1352–1370.

Bollen, N. P. B. & Busse, J. A. (2004), 'Short-term persistence in mutual fund performance', *The Review of Financial Studies* **18**(2), 569–597.

Bollen, N. P. B. & Whaley, R. E. (2009), 'Hedge fund risk dynamics: Implications for performance appraisal', *The Journal of Finance* **64**(2), 985–1035.

Busse, J. A. & Irvine, P. J. (2006), 'Bayesian alphas and mutual fund persistence', *The Journal of Finance* **61**(5), 2251–2288.

Carhart, M. M. (1997), 'On persistence in mutual fund performance', *The Journal of Finance* **52**(1), pp. 57–82.

Chan, J. C. & Koop, G. (2014), 'Modelling breaks and clusters in the steady states of macroeconomic variables', *Computational Statistical and Data Analysis* **76**, 186–193.

Chib, S. (1996), 'Calculating posterior distributions and modal estimates in Markov mixture models', *Journal of Econometrics* **75**, 79–97.

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

Chib, S. & Hamilton, B. (2002), 'Semiparametric Bayes analysis of longitudinal data treatment models', *Journal of Econometrics* **110**, 67–89.

Chib, S. & Tiwari, R. (1988), 'Bayes prediction density and regression estimation: A semi parametric approach', *Empirical Economics* **13**, 209–222.

Dickey, J. D. (1971), 'The weighted likelihood ratio, linear hypotheses on normal location parameters', *Annals of Mathematical Statistics* **42**, 204–223.

Dubey, S. D. (1970), 'Compound gamma, beta and f distributions', *Metrika* **16**(1), 27–31.

Dufays, A. (2015), 'Infinite-state Markov-switching for dynamic volatility', *Journal of Financial Econometrics* **14**(2), 418.

Elton, E. J. & Gruber, M. J. (2013), Mutual funds, *in* M. H. George M. Constantinides & R. M. Stulz, eds, 'Handbook of the Economics of Finance', Vol. 2, Part B, Elsevier, pp. 1011 – 1061.

Elton, E. J., Gruber, M. J. & Blake, C. R. (1996), 'The persistence of risk-adjusted mutual fund performance', *Journal of Business* **69**(2), 133–157.

Escobar, M. D. & West, M. (1995), 'Bayesian density estimation and inference using mixtures', *Journal of the American Statistical Association* **90**(430), 577–588.

Fama, E. F. & French, K. R. (1993), 'Common risk factors in the returns on stocks and bonds', *Journal of Financial Economics* **33**(1), 3 – 56.

Ferguson, T. (1973), 'A Bayesian analysis of some nonparametric problems', *The Annals of Statistics* **1**(2), 209–230.

Fisher, M. (2017), Nonparametric density estimation using a mixture of order-statstics distributions, Technical report, Federal Reserve Bank of Atlanta.
URL: *http://www.markfisher.net/ mefisher/papers/multi-resolution.pdf*

Frühwirth-Schnatter, S. (2006), *Finite Mixture and Markov Switching Models*, Springer.

Frühwirth-Schnatter, S. & Kaufmann, S. (2006), 'How do changes in monetary policy affect bank lending? An analysis of Austrian bank data', *Journal of Applied Econometrics* **21**(3), 275–305.

Frühwirth-Schnatter, S. & Kaufmann, S. (2008), 'Model-based clustering of multiple time series', *Journal of Business & Economic Statistics* **26**(1), 78–89.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A. & Rubin, D. B. (2013), *Bayesian Data Analysis*, Chapman and Hall/CRC.

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

Geweke, J. (2007), 'Interpretation and inference in mixture models: Simple MCMC works', *Computational Statistics & Data Analysis* **51**(7), 3529 – 3550.

Geweke, J. & Jiang, Y. (2011), 'Inference and prediction in a multiple-structual-break model', *Journal of Econometrics* **163**, 172–185.

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

Goetzmann, W. N. & Ibbotson, R. G. (1994), 'Do winners repeat? Patterns in mutual fund return behavior', *Journal of Portfolio Management* **20**(2), 9–18.

Grinblatt, M. & Sheridan, T. (1992), 'The persistence of mutual fund performance', *The Journal of Finance* **47**(5), 1977–1984.

Hendricks, D., Patel, J. & Zeckhauser, R. (1993), 'Hot hands in mutual funds: Short-run persistence of relative performanc, 1974 – 1988', *The Journal of Finance* **48**(1), 93–130.

Hirano, K. (2002), 'Semiparametric Bayesian inference in autoregressive panel data models', *Econometrica* **70**, 781–799.

Hjort, N. L., Holmes, C., Mueller, P. & Walter, S. G., eds (2010), *Bayesian Nonnparametrics*, Cambridge University Press.

Jensen, M. C. (1968), 'The performance of mutual funds in the period 1945-1964', *The Journal of Finance* **23**(2), pp. 389–416.

Jensen, M. J. & Maheu, J. M. (2010), 'Bayesian semiparametric stochastic volatility modeling', *Journal of Econometrics* **157**(2), 306 – 316.

Jin, X. & Maheu, J. M. (2016), 'Bayesian semiparametric modeling of realized covariance matrices', *Journal of Econometrics* **192**(1), 19 – 39.

Jochmann, M. (2015), 'Modeling U.S. inflation dynamics: A Bayesian nonparametric approach', *Econometric Reviews* **34**(5), 537–558.

Jones, C. S. & Shanken, J. (2005), 'Mutual fund performance with learning across funds', *Journal of Financial Economics* **78**, 507–552.

Koijen, R. S. (2014), 'The cross-section of managerial ability, incentives, and risk preferences', *The Journal of Finance* **69**(3), 1051–1098.

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

Lijoi, A., Prünster, I. & Walker, S. G. (2005), 'On consistency of nonparametric normal mixtures for Bayesian density estimation', *Journal of the American Statistical Association* **100**(472), 1292–1296.

MacEachern, S. N. & Müller, P. (1998), 'Estimating mixture of Dirichlet process models', *Journal of Computational and Graphical Statistics* **7**(2), 223–238.

Maheu, J. M. & Song, Y. (2014), 'A new structural break model, with an application to Canadian inflation forecasting', *International Journal of Forecasting* **30**(1), 144 – 160.

Malsiner-Walli, G., Frühwirth-Schnatter, S. & Grün, B. (2016), 'Model-based clustering based on sparse finite Gaussian mixtures', *Statistics and Computing* **26**(1), 303–324.

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

Neal, R. (2000), 'Markov chain sampling methods for Dirichlet process mixture models', *Journal of Computational and Graphical Statistics* **9**, 249–265.

Pesaran, M. H., Pettenuzzo, D. & Timmermann, A. (2006), 'Forecasting time series subject to multiple structural breaks', *Review of Economic Studies* **73**(4), 1057 – 1084.

Petrone, S. & Wasserman, L. (2002), 'Consistency of Bernstein polynomial posteriors', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **64**(1), 79–100.

Song, Y. (2014), 'Modelling regime switching and structural breaks with an infinite hidden Markov model', *Journal of Applied Econometrics* **29**(5), 825–842.

Teh, W. Y. (2010), Dirichlet process, *in* C. Sammut & G. I. Webb, eds, 'Encyclopedia of Machine Learning', Springer US, Boston, MA, chapter Dirichlet Process, pp. 280–287.

West, M., Muller, P. & Escobar, M. (1994), Hierarchical priors and mixture models with applications in regression and density estimation, *in* P. R. Freeman & A. F. Smith, eds, 'Aspects of Uncertainty', John Wiley.

Online Appendix

A The model

In this section we describe a factor model of returns for *fund-regimes*. Within a single fund-regime, the model is quite standard. The novelty lies in two places: the determination of the regimes within a single fund and the connections across the fund-regimes via hyperparameters in the prior distribution.

Fund returns and regimes

Let J denote the number of mutual funds and let $i \in \{1, \ldots, J\}$ index the funds. For fund i there are $\mathcal{T}_i$ observations that occur at times

$$t \in \mathcal{T}_i = \{\tau_i, \tau_i + 1, \ldots, T_i\}, \tag{A.1}$$

where $1 \leq \tau_i < T_i \leq T_{\max}$. Note $\mathcal{T}_i = |\mathcal{T}_i|$. Let $y_{i,t}$ denote the t-th observation for the i-th fund. Let

$$Y_{i,t} := (y_{i\tau_i}, \ldots, y_{it}), \tag{A.2}$$

so that Y_{i,T_i} denotes the full set of returns for fund i.

Let M_i denote the maximum number of regimes for fund i and let $m \in \{1, \ldots, M_i\}$ index the regime number. Let $s_{i,t}$ denote the regime number for fund i at observation t. Note $s_{i,T_i} \leq \min[\mathcal{T}_i, M_i]$. Let

$$S_{i,t} := (s_{i,\tau_i}, \ldots, s_{i,t}), \tag{A.3}$$

so that S_{i,T_i} denotes the full set of regime indicators for fund i. Let $\mathcal{S} := \{S_{i,T_i}\}_{i=1}^{J}$ denote the complete set of states for all funds. We make the following independence assumption regarding the distribution of the observations given the states:

$$p(\{Y_{i,T_i}\}_{i=1}^{J} | \mathcal{S}) = \prod_{i=1}^{J} p(Y_{i,T_i} | S_{i,T_i}) = \prod_{i=1}^{J} \prod_{t=\tau_i}^{T_i} p(y_{i,t} | s_{i,t}). \tag{A.4}$$

Assigned and unassigned regimes

The index set of potential regimes is

$$\mathfrak{I}_{\mathbb{M}} := \{(i, m) : i \in \{1, \ldots, J\} \ \wedge \ m \in \{1, \ldots, M_i\}\}. \tag{A.5}$$

The number of potential regimes is $\mathcal{N}_{\mathcal{M}} = \sum_{i=1}^{J} M_i$. Conditioning on the states, the index set of *assigned* regimes (i.e., those regimes associated with observations) is

$$\mathcal{I}_{\mathcal{A}} := \{(i,m) \in \mathcal{I}_{\mathcal{M}} : m \leq s_{i,T_i}\}, \tag{A.6}$$

which has $\mathcal{N}_{\mathcal{A}} = \sum_{i=1}^{J} s_{i,T_i}$ elements. The index set of unassigned regimes is

$$\mathcal{I}_{\mathcal{U}} := \{(i,m) \in \mathcal{I}_{\mathcal{M}} : m > s_{i,T_i}\} = \mathcal{I}_{\mathcal{M}} \setminus \mathcal{I}_{\mathcal{A}}, \tag{A.7}$$

which has $\mathcal{N}_{\mathcal{U}} = \mathcal{N}_{\mathcal{M}} - \mathcal{N}_{\mathcal{A}}$ elements.[23]

A single fund-regime

The model is connected to the data via the fund-regimes. Let

$$\mathcal{Y}_{i,m} := \{y_{i,t} \in Y_{i,T_i} : s_{i,t} = m\} \tag{A.8}$$

denote the set of observations for regime m in fund i and let $\mathcal{T}_{i,m}$ denote the number of observations in $\mathcal{Y}_{i,m}$. The fund-regime constitutes the unit of analysis. For unassigned regimes, where $m > s_{n,T_n}$, we have $\mathcal{Y}_{i,m} = \varnothing$ and $\mathcal{T}_{i,m} = 0$.

We restrict ourselves to factor models of the following sort. Let N denote the number factors in the model of returns. For $y_{i,t} \in \mathcal{Y}_{i,m}$ given $m \leq s_{i,T_i}$,

$$y_{i,t} = \alpha_{i,m} + \sum_{j=1}^{N} \beta_{i,m}^{j} F_{it}^{j} + \varepsilon_{i,t}, \tag{A.9a}$$

where

$$\varepsilon_{i,t} \overset{\text{iid}}{\sim} \mathsf{N}(0, \sigma_{i,m}^2). \tag{A.9b}$$

For more compact notation, let

$$\phi_{im} := (\phi_{im}^0, \phi_{im}^1, \ldots, \phi_{im}^N) = (\alpha_{i,m}, \beta_{i,m}^1, \ldots, \beta_{i,m}^N) \tag{A.10a}$$

$$\theta_{im} := (\phi_{im}, \sigma_{i,m}^2). \tag{A.10b}$$

In addition, let $X_{it} = (1, F_{it}^1, \ldots, F_{it}^N)$, where $F_{it} = (F_{it}^1, \ldots, F_{it}^N)$ is the vector of market-wide factors properly-aligned to fund i. With this notation, we can express (A.9) as

$$p(y_{i,t}|s_{i,t} = m) = p(y_{i,t}|\theta_{im}) = \mathsf{N}(y_{i,t}|X_{it}^{\top}\phi_{im}, \sigma_{i,m}^2). \tag{A.11}$$

[23] One way to eliminate structrual breaks is to set $M_i = 1$ for all i, in which case $\mathcal{I}_{\mathcal{M}} = \mathcal{I}_{\mathcal{A}}$, $\mathcal{I}_{\mathcal{U}} = \varnothing$, and $\mathcal{N}_{\mathcal{M}} = J$. Another way is to set the probability of a regime change to zero for all funds. See below.

Consequently, the likelihood for θ_{im} can be expressed as

$$p(\mathcal{Y}_{im}|\theta_{im}) = \mathsf{N}(\mathcal{Y}_{im}|\mathcal{X}_{im}\,\phi_{im}, \sigma^2_{i,m} I_{\mathcal{T}_{im}}), \tag{A.12}$$

where

$$\mathcal{X}_{im} := \{X_{it} : s_{i,t} = m\}. \tag{A.13}$$

We complete the model of a fund-regime by providing a prior for the fund-regime parameters. The prior for θ_{im} (conditional on hyperparameters) is[24]

$$p(\theta_{im}|a_{im}, h^2_{\phi im}, h^2_{\sigma im}) = \mathsf{N}(\phi_{im}|a_{im}, B_{im})\,\textsf{Inv-Gamma}(\sigma^2_{im}|\nu_\sigma/2, h^2_{\sigma im}\,\nu_\sigma/2), \tag{A.14}$$

where $B_{im} = \text{diag}(h^2_{\phi im})$ is a $(N+1) \times (N+1)$ diagonal matrix and

$$a_{im} := (a^0_{im}, a^1_{im}, \ldots, a^N_{im}) \tag{A.15a}$$

$$h^2_{\phi im} := (h^{2,0}_{\phi im}, h^{2,1}_{\phi im}, \ldots, h^{2,N}_{\phi im}). \tag{A.15b}$$

Reexpressing the hyperparameters

Before we proceed, it is convenient to express the hyperparameters in the prior for θ_{im} [see (A.15)] as follows. Let

$$\psi^j_{im} := (a^j_{im}, h^{2,j}_{\phi im}) \qquad \text{and} \qquad \psi^\sigma_{im} := h^2_{\sigma im}. \tag{A.16}$$

Using this notation, we can reexpress the prior for θ_{im} [see (A.14)] as

$$\phi^j_{im}|\psi^j_{im} \sim \mathsf{N}(a^j_{im}, h^{2,j}_{\phi im}) \qquad \text{for } j = 0, \ldots, N \tag{A.17a}$$

$$\sigma^2_{im}|\psi^\sigma_{im} \sim \textsf{Inv-Gamma}(\nu_\sigma/2, h^2_{\sigma im}\,\nu_\sigma/2). \tag{A.17b}$$

This formulation will prove useful in computing the posterior distributions for the hyperparameters.

[24] For reference, note

$$\textsf{Inv-Gamma}(\sigma^2|\nu/2, s^2\,\nu/2) = \frac{e^{-\frac{\nu s^2}{2\,\sigma^2}} \left(\frac{\nu s^2}{2\,\sigma^2}\right)^{\nu/2}}{s^2\,\Gamma\left(\frac{\nu}{2}\right)}$$

$$\textsf{Gamma}(s^2|a, 1/b) = \frac{e^{-b\,s^2}\,(b\,s^2)^{a-1}}{b^{-1}\,\Gamma(a)}.$$

Prior for the regimes

We complete the model of a fund by providing a prior for the regime changes within a fund.

Let q_i denote the probability of a change in regime (i.e., state) for the i-th fund. The conditional prior for the regimes within a fund is

$$p(S_{iT_i}|q_i) = p(s_{i,\tau_i}) \prod_{t=\tau_i}^{T_i-1} p(s_{i,t+1}|s_{it}, q_i), \tag{A.18}$$

where $p(s_{i,\tau_i} = 1) = 1$ and

$$p(s_{i,t+1} = \ell | s_{i,t} = m, q_i) = \begin{cases} 1 - q_i & \ell = m < M_i \\ q_i & \ell = m + 1 \\ 1 & \ell = m = M_i \\ 0 & \text{otherwise} \end{cases}. \tag{A.19}$$

The prior for q_i is given by

$$p(q_i|\psi_i^q) = \mathsf{Beta}(q_i|j_i, k_i - j_i + 1), \tag{A.20}$$

where

$$\psi_i^q := (j_i, k_i) \tag{A.21}$$

denotes the hyperparameter in the prior for q_i. We assume conditional independence across funds as follows:

$$p\left(\{(S_{iT_i}, q_i)\}_{n=1}^J \,\middle|\, \{\psi_i^q\}_{i=1}^J\right) = \prod_{i=1}^J p(S_{iT_i}|q_i)\, p(q_i|\psi_i^q). \tag{A.22}$$

Priors for the hyperparameters

A central feature of the model is the sharing of hyperparameters across fund-regimes. The mechanism for sharing involves grouping hyperparameters into clusters and assigning a common value to all hyperparameters within a given cluster — the value of the *cluster parameter*. In other words, we identify a fund-regime parameter with the parameter of the cluster to which it has been assigned. We adopt the Chinese Restaurant Process (CRP) representation of the Dirichlet process prior distribution for the clustering. The cluster parameters themselves have a prior distribution called the *base distribution* (which the fund-regime hyperparameters inherit).

We first provide a skeleton for the hyperprior that can be applied to the various cases.

Hyperprior skeleton

We begin with the classifications. The CRP can be characterized as follows:

$$z_{1:K}|\eta \sim \mathsf{CRP}(\eta), \tag{A.23}$$

where $z_{1:K}$ denotes a collection of the K classifications variables, z_κ, $\kappa = 1, \ldots, K$, and η denotes the concentration parameter of the Dirichlet process $DP(\eta, G_0)$. In order to provide an explicit representation, let $\mathcal{K}_\kappa = \max[z_{1:\kappa}]$, the number of clusters in $z_{1:\kappa}$, and let n_c^κ denote the multiplicity of c in $z_{1:\kappa}$ so that $\sum_{c=1}^{\kappa} n_c^\kappa = \kappa$. Then (A.23) can be expressed as

$$p(z_{1:\mathcal{K}}|\eta) = \prod_{\kappa=1}^{K-1} p(z_{\kappa+1}|z_{1:\kappa}, \eta), \tag{A.24}$$

where $z_1 = 1$ and

$$p(z_{\kappa+1} = c|z_{1:\kappa}, \eta) = \begin{cases} \frac{n_c^\kappa}{\kappa+\eta} & c \in \{1, \ldots, \mathcal{K}_\kappa\} \\ \frac{\eta}{\kappa+\eta} & c = \mathcal{K}_\kappa + 1 \end{cases}. \tag{A.25}$$

Having determined the classifications, a cluster parameter is drawn from the base distribution G_0 for each of the $\mathcal{K}_K$ clusters:

$$\psi_c^* \overset{\text{iid}}{\sim} G_0. \tag{A.26}$$

Finally, combining the classifications with the cluster parameters, a hyperparameter is set equal to the cluster parameter for the cluster to which it (the hyperparameter) belongs:

$$\psi_\kappa = \psi_{z_\kappa}^*. \tag{A.27}$$

Details for clustering

We put a prior on all of the hyperparameters associated with the full set of potential factor-model coefficients; namely,

$$\{\psi_{im}^j\}_{\mathcal{M}} := \{\psi_{im}^j : (i,m) \in \mathcal{I}_{\mathcal{M}}\} \tag{A.28a}$$

$$\{\psi_{im}^\sigma\}_{\mathcal{M}} := \{\psi_{im}^\sigma : (i,m) \in \mathcal{I}_{\mathcal{M}}\}. \tag{A.28b}$$

Let z_{im}^j denote the cluster to which ψ_{im}^j is assigned, so that $z_{im}^j = c$ means $\psi_{im}^j = \psi_c^{j*}$. Similarly, let z_{im}^σ denote the cluster assignment for ψ_{im}^σ. The prior for the cluster assignments can be expressed as

$$\{z_{im}^j\}_{\mathcal{M}}|\eta_j \sim \mathsf{CRP}(\eta_j) \tag{A.29a}$$

$$\{z_{im}^\sigma\}_{\mathcal{M}}|\eta_\sigma \sim \mathsf{CRP}(\eta_\sigma). \tag{A.29b}$$

where η_j and η_σ are the DP prior's *concentration parameters*.

Let z_i^q denote the cluster assignment for ψ_i^q and let $\{z_i^q\}_{1:J}$ denote the set of assignments for all funds. The prior for the classifications can be expressed as

$$\{z_i^q\}_{1:J}|\eta_q \sim \mathsf{CRP}(\eta_q), \tag{A.30}$$

where η_q is the concentration parameter.

The prior for the concentration parameters is given by

$$\log(\eta_\ell) \sim \mathsf{Logistic}(1,1), \tag{A.31}$$

for $\ell \in \{j, \sigma, q\}$.[25]

Base distributions

We now turn to the base distributions for the cluster parameters. Let us denote the cluster parameters by

$$\psi_c^{j*} = (a_c^{j*}, h_{\phi c}^{2j*}), \qquad \psi_c^{\sigma*} = h_{\sigma c}^{2*}, \qquad \text{and} \qquad \psi_c^{q*} = (j_c^*, k_c^*), \tag{A.32}$$

where c is an index for the cluster number. (Each of the cluster parameters has its own index c.) The base distributions for ψ_c^{j*} and $\psi_c^{\sigma*}$ are given by

$$p(\psi_c^{j*}) = p(a_c^{j*}, h_{\phi c}^{2j*}) = \mathsf{N}(a_c^{j*}|a_0^j, h_{\phi c}^{2j*}/\kappa_0^j)\,\mathsf{Inv\text{-}Gamma}(h_{\phi c}^{2j*}|\nu_0^j/2, h_0^{2j}\,\nu_0^j/2) \tag{A.33a}$$

and

$$p(\psi_c^{\sigma*}) = p(h_{\sigma c}^{2*}) = \mathsf{Gamma}(h_{\sigma c}^{2*}|c_0, 1/b_0). \tag{A.33b}$$

The base distribution for ψ_c^{q*} can be expressed as $p(\psi_c^{q*}) = p(j_c^*, k_c^*) = p(j_c^*|k_c^*)\,p(k_c^*)$, where[26]

$$j_c^*|k_c^* \sim \mathsf{Uniform}(1, \ldots, k_c^*) \tag{A.34a}$$

$$k_c^* - 1 \sim \mathsf{Geometric}(\xi_0). \tag{A.34b}$$

[25] For reference, note $\mathsf{Log\text{-}Logistic}(x|a,b) = ab^{-a}x^{a-1}/(1+(x/b)^a)^2$.

[26] Let $\xi_0 = 1/200$.

B Sampling from the posterior

We now present our approach to sampling from the posterior distribution. We divide the parameters into blocks and sample each block conditional on the parameters in the other blocks (Gibbs sampling approach). For each block we adopt an appropriate strategy for sampling.

There are two types of blocks: within-fund blocks and across-fund blocks. The within-fund blocks are conditionally independent across funds and consequently can be handled in parallel. The across-fund blocks involve drawing the hyperparameters (via the classifications and the cluster parameters) and the concentration parameters.

Fund-regime parameters $\{\theta_{im}\}$

Given the likelihood (A.12) and the conditionally conjugate prior (A.14), the two conditional posteriors are given by

$$p(\phi_{im}|\mathcal{Y}_{im}, \sigma^2_{im}, a_{im}, B_{im}) = \mathsf{N}(\phi_{im}|\overline{a}_{im}, \overline{B}_{im}) \tag{B.1a}$$

where

$$\overline{B}_{im} = \left(\sigma^{-2}_{im}\mathcal{X}^\top_{im}\mathcal{X}_{im} + B^{-1}_{im}\right)^{-1} \tag{B.1b}$$

$$\overline{a}_{im} = \overline{B}_{im}\left(\sigma^{-2}_{im}\mathcal{X}^\top_{im}\mathcal{Y}_{im} + B^{-1}_{im}\, a_{in}\right) \tag{B.1c}$$

and

$$p(\sigma^2_{im}|\mathcal{Y}_{nm}, \phi_{im}, h^2_{\sigma im}) = \mathsf{Inv\text{-}Gamma}(\sigma^2_{im}|\overline{\nu}_{im}/2, \overline{h}^2_{\sigma im}\overline{\nu}_{im}/2) \tag{B.2a}$$

where

$$\overline{\nu}_{im} = \nu_\sigma + \mathcal{T}_{im} \tag{B.2b}$$

$$\overline{h}^2_{\sigma im} = \frac{h^2_{\sigma im}\,\nu_\sigma + (\mathcal{Y}_{im} - \mathcal{X}_{im}\phi_{im})^\top(\mathcal{Y}_{im} - \mathcal{X}_{im}\phi_{im})}{\overline{\nu}_{im}}. \tag{B.2c}$$

For unassigned regimes (those regimes for which $\mathcal{Y}_{im} = \varnothing$), the posterior reduces to the prior (conditional on the hyperparameters). [See (A.14).]

Regimes $\{S_{iT_i}\}$

Factor the joint posterior distribution of the states as follows:

$$\begin{aligned} p(S_{iT_i}|Y_{i,T_i}, \Theta_i, q_i) = p(s_{i,T_i}|Y_{i,T_i}, \Theta_i, q_i)\, p(s_{i,T_i-1}|S^{T_i}_i, Y_{i,T_i}, \Theta_i, q_i)\ \cdots \\ p(s_{i,t}|S^{t+1}_i, Y_{i,T_i}, \Theta_i, q_i)\ \cdots\ p(s_{i,1}|S^2_i, Y_{iT_i}, \Theta_i, q_i), \end{aligned} \tag{B.3}$$

where

$$\Theta_i = (\theta_{i1}, \ldots, \theta_{iM_i}) \tag{B.4a}$$

$$S_i^t = (s_{i,t}, \ldots, s_{i,T_i}). \tag{B.4b}$$

Draws of the states will be made using (B.3) starting with s_{i,T_i} and working backwards.

The conditional distributions in (B.3) can be computed as follows:

$$p(s_{i,t}|S_i^{t+1}, Y_{i,T_i}, \Theta_i, q_i) \propto p(s_{i,t}|Y_{i,t}, \Theta_i, q_i)\, p(s_{i,t+1}|s_{i,t}, q_i), \tag{B.5}$$

where

$$p(s_{i,t} = m|Y_{it}, \Theta_i, q_i) = \frac{p(s_{i,t} = m|Y_{i,t-1}, \Theta_i, q_i)\, p(y_{i,t}|\theta_{im})}{\sum_{m'=1}^{M_i} p(s_{it} = m'|Y_{i,t-1}, \Theta_i, q_i)\, p(y_{i,t}|\theta_{i,m'})} \tag{B.6}$$

and where

$$\begin{aligned} &p(s_{i,t} = m|Y_{i,t-1}, \Theta_i, q_i) \\ &\qquad = \sum_{m'=m-1}^{m} p(s_{i,t} = m|s_{i,t-1} = m', q_i)\, p(s_{i,t-1} = m'|Y_{i,t-1}, \Theta_i, q_i) \end{aligned} \tag{B.7}$$

for $t \geq 2$ and $p(s_{i,1} = 1|Y_{i0}, \Theta_i, q_i) = 1$.

Let us streamline the notation and bring out the recursive structure. We define three matrices. First, let Q_i denote the $M_i \times M_i$ matrix where [see (A.19)]

$$(Q_i)_{m\ell} = p(s_{i,t+1} = \ell|s_{i,t} = m, q_i). \tag{B.8}$$

Second, let L^i denote the $T_i \times M_i$ matrix where

$$L^i_{tm} = p(y_{i,t}|\theta_{im}) = \mathsf{N}(y_{i,t}|X_{it}^\top \phi_{im}, \sigma^2_{i,m}), \tag{B.9}$$

and let L^i_t denote the tth row of L^i. Third, let G^i denote the $T_i \times M_i$ matrix where[27]

$$G^i_{tm} = p(s_{i,t} = m|Y_{i,t-1}, \Theta_i, q_i)\, p(y_{i,t}|\theta_{im}) \tag{B.10}$$

and similarly let G^i_t denote the tth row of G^i.

With this notation, the forward recursion is given by

$$G^i_t \propto (G^i_{t-1} Q_i) \circ L^i_t, \tag{B.11}$$

[27] Note $G^i_{tm} = p(s_{i,t} = m, y_{i,t}|Y_{i,t-1}, \Theta_i, q_i)$.

where "$\circ$" denotes the Hadamard (component-by-component) product. The recursion is initialized with $G_1^i = (L_{11}^i, 0, \ldots, 0)$. The probabilities of the states at $t = T_i$ are given by

$$p(s_{i,T_i} = m | Y_{i,T_i}, \Theta_i, q_i) \propto G_{T_i m}^i, \tag{B.12}$$

while for $t < T_i$ they are

$$p(s_{i,t} = m | s_{i,t+1} = \ell, Y_{i,T_i}, \Theta_i, q_i) \propto (G_t^i \circ Q_i^\ell)_m, \tag{B.13}$$

where Q_i^ℓ denotes the ℓth column of Q_i.[28]

The backward sampler begins with (B.12) and continues with (B.13) iteratively until $s_{i,1}$ is drawn.

Regime-change probabilities $\{q_i\}$

Given the states and the classifications, we can draw q_i as follows. The likelihood for q_i is

$$\mathsf{Binomial}(s_{i,T_i} - 1 | \mathcal{T}_i - 1, q_i), \tag{B.14}$$

since $s_{iT_i} - 1$ is the number of change-points in $T_i - 1$ independent Bernoulli trials where q_i is the probability of a change-point for a single Bernoulli trial. Therefore the conditional posterior distribution is

$$p(q_i | s_{i,T_i}, \psi_i^q) = \mathsf{Beta}\big(q_i | j_i + s_{i,T_i} - 1,\, k_i - j_i + 1 + (\mathcal{T}_i - s_{i,T_i})\big). \tag{B.15}$$

Hyperparameters

We now turn to drawing the hyperparameters. Recall there are three sets of hyperparameters: $\{\psi_{im}^j\}_{\mathcal{M}}$, $j = 0, 1, \ldots, N$, $\{\psi_{im}^\sigma\}_{\mathcal{M}}$, and $\{\psi_i^q\}_{1:J}$, where $\psi_{im}^j = (a_{im}^j, h_{\phi im}^{2j})$, $\psi_{im}^\sigma = h_{\sigma im}^2$, and $\psi_i^q = (j_i, k_i)$.

Regarding ψ_{im}^j and ψ_{im}^σ, it is necessary to distinguish between those hyperparameters associated with regimes that have been assiged (conditional on the states $\mathcal{S}$) and those that have not. In particular, let $\{\psi_{im}^j\}_{\mathcal{A}}$ and $\{\psi_{im}^\sigma\}_{\mathcal{A}}$ denote the sets of these hyperparameters for regimes that have been assigned and let $\{\psi_{im}^j\}_{\mathcal{U}}$ and $\{\psi_{im}^\sigma\}_{\mathcal{U}}$ denote the sets of these hyperparameters for regimes that are unassigned.

For each of $\{\psi_{im}^j\}_{\mathcal{A}}$, $\{\psi_{im}^\sigma\}_{\mathcal{A}}$, and $\{\psi_i^q\}_{1:J}$, we adopt the approach to sampling based on Algorithm 2 in Neal (2000), although the details differ across the three sets of hyperparameters. For the remaining two sets of hyperparameters, $\{\psi_{im}^j\}_{\mathcal{U}}$ and $\{\psi_{im}^\sigma\}_{\mathcal{U}}$, we sample from the distributions conditioned on the respective assigned hyperparameters $\{\psi_{im}^j\}_{\mathcal{A}}$ and $\{\psi_{im}^\sigma\}_{\mathcal{A}}$.

[28] If $s_{i,t} = 1$ then $s_{i,t-1} = 1$, while if $s_{i,t} = t$, then $s_{i,t-1} = t - 1$. In either of these cases no more computation is required.

Drawing $\{\psi^j_{im}\}_{\mathcal{A}}$

We must establish some additional notation. Define the set of indices of specific regimes associated with cluster c,

$$\mathcal{I}^j_c := \{(i,m) \in \mathcal{I}_{\mathcal{A}} : z^j_{im} = c\}, \tag{B.16}$$

and the set of parameters associated with cluster c,

$$\{\phi^j_{im}\}_c := \{\phi^j_{im} | (i,m) \in \mathcal{I}^j_c\}. \tag{B.17}$$

Let n^j_c denote the number of elements in $\{\phi^j_{im}\}_c$. Let $\{z^j_{im}\}_{\mathcal{A}}$ denote the set of classifications for assigned regimes. Assume $\{z^j_{im}\}_{\mathcal{A}}$ is normalized so that the set of unique values equals $\{1, 2, \ldots, \mathcal{K}_j\}$, where $\mathcal{K}_j$ is the number of distinct classifications. Note $\sum_{c=1}^{\mathcal{K}_j} n^j_c = \mathcal{N}_{\mathcal{A}}$.

Note that $\{\phi^j_{im}\}_c$ plays the role of the "observations" with regard to the "parameter" ψ^{j*}_c in the conditional posterior. In this role, the likelihood for a single observation is given by the prior (A.17a). Thus the posterior for ψ^{j*}_c given $\{\phi^j_{im}\}_c$ is:

$$p(\psi^{j*}_c | \{\phi^j_{im}\}_c) = \mathsf{N}(a^{j*}_c | \overline{a}_j, h^{2j*}_{\phi c}/\overline{\kappa}_j)\, \mathsf{Inv\text{-}Gamma}(h^{2j*}_{\phi c} | \overline{\nu}_j/2, \overline{h}^2_j\, \overline{\nu}_j/2), \tag{B.18}$$

where

$$\overline{\kappa}_j = \kappa^j_0 + n^j_c \tag{B.19a}$$

$$\overline{\nu}_j = \nu^j_0 + n^j_c \tag{B.19b}$$

$$\overline{a}_j = \left(\frac{\kappa^j_0}{\overline{\kappa}_j}\right) a^j_0 + \left(\frac{n^j_c}{\overline{\kappa}_j}\right) \overline{\phi}^j_c \tag{B.19c}$$

$$\overline{h}^2_j = \left(\frac{\nu^j_0}{\overline{\nu}_j}\right) h^{j2}_0 + \left(\frac{n^j_c}{\overline{\nu}^j_0}\right) \widehat{\sigma}^2_{\phi^j_c} + \left(\frac{\kappa^j_0}{\overline{\kappa}_j}\right)\left(\frac{n^j_c}{\overline{\nu}_j}\right) (\overline{\phi}^j_c - a^j_0)^2, \tag{B.19d}$$

where the "observations" are summarized by

$$\overline{\phi}^j_c := \frac{1}{n^j_c} \sum_{\iota \in \mathcal{I}^j_c} \phi^j_\iota \qquad \text{and} \qquad \widehat{\sigma}^2_{\phi^j_c} := \frac{1}{n^j_c} \sum_{\iota \in \mathcal{I}^j_c} (\phi^j_\iota - \overline{\phi}^j_c)^2. \tag{B.20}$$

Classification

It remains to describe how the elements of $\{z^j_{im}\}_{\mathcal{A}}$ (the classifications for the assigned regimes) are drawn. Let $\{z^j_{im}\}^{-\iota}_{\mathcal{A}}$ denote the (possibly renormalized) vector of classifications after having removed case ι for some $\iota \in \mathcal{I}_{\mathcal{S}}$. Renormalization will occur if — before removal — the cluster associated with observation ι is a singleton (i.e., $n^j_{z^j_\iota} = 1$), in which case $\psi^{j*}_{z^j_\iota}$

will be discarded, the remaining clusters will be relabeled, and the corresponding classifications will be adjusted. Let $\mathcal{K}_j^{-\iota} = \max[\{z_{im}^j\}_{\mathcal{A}}^{-\iota}]$ and let $(n_c^j)^{-\iota}$ denote the multiplicity of class c in $\{z_{im}^j\}_{\mathcal{A}}^{-\iota}$.

The full conditional probability of z_ι^j given the "observations" is

$$p(z_\iota^j | \{z_{im}^j\}_{\mathcal{A}}^{-\iota}, \phi_\iota^j, \eta_j) \propto \begin{cases} \frac{(n_c^j)^{-\iota}}{\mathrm{N}_{\mathcal{A}} - 1 + \eta_j} \, p(\phi_\iota^j | z_\iota^j = c) & c \in \{1, \ldots, \mathcal{K}_j^{-\iota}\} \\ \frac{\eta_j}{\mathrm{N}_{\mathcal{A}} - 1 + \eta_j} \, p(\phi_\iota^j) & c = \mathcal{K}_j^{-\iota} + 1 \end{cases}, \tag{B.21}$$

where

$$p(\phi_{im}^j | z_{im}^j = c) = \mathsf{N}(\phi_{im}^j | a_c^{j*}, h_{\phi c}^{2j*}), \tag{B.22}$$

and

$$\begin{aligned} p(\phi_{im}^j) &= \int \mathsf{N}(\phi_{im}^j | a_c^{j*}, h_{\phi c}^{2j*}) \, \mathsf{N}(a_c^{j*} | a_0^j, h_{\phi c}^{2j*} / \kappa_0^j) \, \mathsf{Inv\text{-}Gamma}(h_{\phi c}^{2j*} | \nu_0^j / 2, h_0^{j2} \, \nu_0^j / 2) \, d\psi_c^{j*} \\ &= t_{\nu_0^j}(\phi_{im}^j | a_0^j, (1 + 1/\kappa_0^j) \, h_0^{j2}). \end{aligned} \tag{B.23}$$

If a new cluster is chosen, then populate it with a draw from the posterior $\psi_c^{j*} | \phi_{im}^j$ [see (B.18)].

Drawing $\{\psi_{im}^\sigma\}_{\mathcal{A}}$

Let

$$\mathcal{I}_c^\sigma := \{(i, m) \in \mathcal{I}_{\mathcal{A}} : z_{im}^\sigma = c\}, \tag{B.24}$$

and let $\{\sigma_{im}^2\}_c := \{\sigma_{im}^2 | (i, m) \in \mathcal{I}_c^\sigma\}$, where the number observations in $\{\sigma_{im}^2\}_c$ equals n_c^σ. Let $\mathcal{K}_\sigma$ denote the number of distinct classifications in the collection $\{z_{im}^\sigma\}$.

Again, $\{\sigma_{im}^2\}_c$ play the role of the observations where the likelihood for a single observation is given by (A.17b). The posterior for $h_{\sigma c}^{2*}$ given $\{\sigma_{im}^2\}_c$ is

$$p(h_{\sigma c}^{2*} | \{\sigma_{im}^2\}_c) = \mathsf{Gamma}(h_{\sigma c}^{2*} | \bar{c}_c, 1/\bar{b}_c), \tag{B.25}$$

where

$$\bar{c}_c = c_0 + \frac{\nu_\sigma}{2} \, n_c^\sigma \tag{B.26a}$$

$$\bar{b}_c = b_c + \frac{\nu_\sigma}{2} \sum_{\iota \in \mathcal{I}_c^\sigma} \frac{1}{\sigma_\iota^2}. \tag{B.26b}$$

Classification

The probabilities for classification involve the following likelihoods: For existing clusters,

$$p(\sigma^2_{im}|z^\sigma_{im} = c) = \textsf{Inv-Gamma}(\sigma^2_{im}|\nu_\sigma/2, h^{2*}_{\sigma c}\,\nu_\sigma/2), \tag{B.27}$$

while for a new cluster we use the predictive distribution:

$$\begin{aligned} p(\sigma^2_{im}) &= \int \textsf{Inv-Gamma}(\sigma^2_{im}|\nu_\sigma/2, h^{2*}_{\sigma c}\,\nu_\sigma/2)\,\textsf{Gamma}(h^{2*}_{\sigma c}|c_0, 1/b_0)\,dh^{2*}_{\sigma c} \\ &= \frac{\left(\frac{\sigma^2_{im}}{w} + 1\right)^{-(c_0+\frac{\nu_\sigma}{2})}\left(\frac{\sigma^2_{im}}{w}\right)^{c_0-1}}{w\,B\big(c_0, \frac{\nu_\sigma}{2}\big)} \qquad \text{where } w = \nu_\sigma/(2\,b_0), \end{aligned} \tag{B.28}$$

which is the PDF of the three-parameter generalized Beta-Prime distribution.[29] If a new cluster is chosen, then populate it with a draw from the posterior $h^{2*}_{\sigma c}|\sigma^2_{im}$ [see (B.25)].

Unassigned regimes

Regarding $\{\psi^j_{im}\}_{\mathcal{U}}$ and $\{\psi^\sigma_{im}\}_{\mathcal{U}}$, we draw the classifications from the CRP conditioned on the cluster counts of the assigned hyperparameters:

$$\{z^j_{im}\}_{\mathcal{U}}|\{z^j_{im}\}_{\mathcal{A}}, \eta_j \sim \textsf{CRP}(\{n^j_c\}_{c=1}^{\mathcal{K}_j}, \eta_j) \tag{B.29a}$$

$$\{z^\sigma_{im}\}_{\mathcal{U}}|\{z^\sigma_{im}\}_{\mathcal{A}}, \eta_\sigma \sim \textsf{CRP}(\{n^\sigma_c\}_{c=1}^{\mathcal{K}_\sigma}, \eta_\sigma), \tag{B.29b}$$

where $\textsf{CRP}(\{n_c\}_{c=1}^{\mathcal{K}}, \eta)$ indicates conditioning on the existing classifications. The unassigned hyperparameters that are classified with already existing clusters are set equal to the cluster parameters from those clusters. For each new cluster drawn via (B.29), a cluster parameter is drawn from the associated base distribution [see (A.33)].

Drawing $\{\psi^q_i\}_{1:J}$

We adopt the scheme described in Fisher (2017) (see for omitted details), which uses Neal's Algorithm 2.

Let

$$\mathcal{I}^q_c := \{i \in \{1, \ldots, J\} : z^q_i = c\}, \tag{B.30}$$

and let $\{q_i\}_c := \{q_i | i \in \mathcal{I}^q_c\}$, where the number observations in $\{q_i\}_c$ equals n^q_c. Let $\mathcal{K}_q$ denote the number of distinct classifications in the collection $\{z^q_i\}$.

[29] Since $\textsf{Inv-Gamma}(\sigma^2_{im}|\nu_\sigma/2, h^{2*}_{\sigma c}\,\nu_\sigma/2) \equiv \textsf{Gamma}(h^{2*}_{\sigma c}|\nu_\sigma/2 + 1, 2\,\sigma^2_{im}/\nu_\sigma)$, (B.28) is also the PDF of a "compound gamma" distribution.

Fisher (2017) presents two methods for drawing ψ_c^{q*}. The first method assumes $n_c^q = 1$. In this case, the conditional posterior is given by

$$p(j_c^*, k_c^*|q_i) = \frac{p(q_i|j_c^*, k_c^*)\, p(j_c^*|k_c^*)\, p(k_c^*)}{p(q_i)} = p(j_c^*|k_c^*, q_i)\, p(k_c^*), \tag{B.31}$$

where

$$p(j_c^*|k_c^*, q_i) = \binom{k_c^* - 1}{j_c^* - 1} q_i^{j_c^* - 1} (1 - q_i)^{k_c^* - j_c^*} = \mathsf{Binomial}(j_c^* - 1|k_c^* - 1, q_i). \tag{B.32}$$

Note that k_c^* is not identified and is drawn from its prior distribution and then $(j_c^* - 1) \sim \mathsf{Binomial}(k_c^* - 1, q_i)$ with the proviso that if $k_c^* = 1$ then $j_c^* = 1$.

Now assume $n_c^q \geq 2$. The posterior for (j_c^*, k_c^*) is given by

$$\begin{aligned} p(\psi_c^{q*}|\{q_i\}, \{z_i^q\}, \eta) &= p(\psi_c^{q*}|\{q_i\}_c) \\ &= p(j_c^*, k_c^*) \times \frac{\left(\prod_{i \in \mathcal{I}_c^q} q_i\right)^{j_c^* - 1} \left(\prod_{i \in \mathcal{I}_c^q} (1 - q_i)\right)^{k_c^* - j_c^*}}{B(j_c^*, k_c^* - j_c^* + 1)^{n_c^q}}. \end{aligned} \tag{B.33}$$

One can adopt a Metropolis-Hastings scheme for the $(r+1)$ draw given the rth draw with the following proposal:

$$k_c^{*'} - 1 \sim \mathsf{Poisson}(k_c^{*(r)}) \tag{B.34}$$

$$j_c^{*'} - 1 \sim \mathsf{Binomial}(k_c^{*'} - 1, \overline{q}_c^{(r)}), \tag{B.35}$$

where $\overline{q}_c = \frac{1}{n_c^q} \sum_{i \in \mathcal{I}_c^q} q_i$ is the average of $\{q_i\}_c$. Let

$$\widehat{p}((\psi_c^{q*})'|\psi_c^{q*}, \overline{q}_c) := \mathsf{Poisson}(k_c^{*'} - 1|k_c^*)\, \mathsf{Binomial}(j_c^{*'}|k_c^{*'} - 1, \overline{q}_c). \tag{B.36}$$

Then

$$\psi_c^{q*\,(r+1)} = \begin{cases} (\psi_c^{q*})' & \mathcal{M}_c^{(r)} \geq u^{(r+1)}, \\ \psi_c^{q*\,(r)} & \text{otherwise}, \end{cases} \tag{B.37}$$

where $u^{(r+1)} \sim \mathsf{Uniform}(0, 1)$ and

$$\mathcal{M}_c^{(r)} = \frac{p((\psi_c^{q*})'|\{q_i\}_c^{(r)})}{p((\psi_c^{q*})^{(r)}|\{q_c\}_c^{(r)})} \times \frac{\widehat{p}((\psi_c^{q*})^{(r)}|(\psi_c^{q*})', \overline{q}_c^{(r)})}{\widehat{p}((\psi_c^{q*})'|(\psi_c^{q*})^{(r)}, \overline{q}_c^{(r)})}. \tag{B.38}$$

Classification

The full conditional probability of z_i given the "observations" is

$$p(z_i|\eta_q, \{z_i^q\}^{-i}, \{\psi_c^*\}^{-i}, q_i) \propto \begin{cases} \frac{(n_c^q)^{-i}}{J - 1 + \eta_q} \mathsf{Beta}(q_i|j_c^*, k_c^* - j_c^* + 1) & c \in \{1, \ldots, \mathcal{K}_q^{-i}\} \\ \frac{\eta_q}{J - 1 + \eta_q} \mathsf{Uniform}(q_i|0, 1) & c = \mathcal{K}_q^{-i} + 1 \end{cases}. \tag{B.39}$$

The uniform predictive for a new cluster follows from the adding-up property of Bernstein polynomials in conjunction with the uniform conditional prior for $j_i^*|k_i^*$.

If a new cluster is selected, then the new cluster parameter is drawn from the posterior using the scheme described above for $n_i^q = 1$.

Concentration parameters

For the purposes of this section, let η stand for any of the concentration parameters (η_j, η_τ, or η_q). In addition, let $\{z_i\}$ denote the corresponding set of classifications, let n_c denote the number of elements in cluster c, and let $\mathcal{K}$ denote the number of clusters. Let $B = \sum_{c=1}^{\mathcal{K}} n_c$. (Note, $B = \mathcal{N}_{\mathcal{S}}$ for η_j and η_σ while $B = J$ for η_q.) It can be shown that the likelihood for the concentration parameter is

$$p(\{z_i\}|\eta) = \frac{\eta^{\mathcal{K}} \prod_{c=1}^{\mathcal{K}} (n_c - 1)!}{(\eta)_B} \propto \frac{\eta^{\mathcal{K}}}{(\eta)_B}, \tag{B.40}$$

where $(\eta)_B := \prod_{j=1}^{B} (j - 1 + \eta) = \Gamma(B + \eta)/\Gamma(\eta)$.

It is convenient to change variables to $w = \log(\eta)$. Recall $w \sim \mathsf{Logistic}(1, 1)$. We can express the product of the likelihood and the prior in terms of w:

$$p(\{z_i\}|w)\, p(w) = \frac{e^{(\mathcal{K}+1)\, w}\, \Gamma(e^w)}{(1 + e^w)^2\, \Gamma(B + e^w)}. \tag{B.41}$$

Draws of w can be made via a Metropolis scheme using a symmetric proposal distribution such as $w' \sim \mathsf{Logistic}(w^{(r)}, h)$, where $w^{(r)} = \log(\eta^{(r)})$. The next element in the chain is given by

$$\eta^{(r+1)} = \begin{cases} e^{w'} & \frac{p(\{z_i\}|w')\, p(w')}{p(\{z_i\}|w^{(r)})\, p(w^{(r)})} \geq u^{(r+1)}, \\ \eta^{(r)} & \text{otherwise,} \end{cases} \tag{B.42}$$

where $u^{(r+1)} \sim \mathsf{Uniform}(0, 1)$.

C Posterior distributions

In this section we discuss the posterior distributions for the population (generic) and specific mutual fund (specific) cases.

Population distribution of the factor coefficients

Let $\mathbb{Y} := \{Y_{iT_i}\}_{i=1}^{J}$. For each factor coefficient there is a posterior predictive distribution for the generic case of a gth unobserved mutual fund:

$$p(\phi_g^j|\mathbb{Y}) = \int p(\phi_g^j|\vartheta_j)\, p(\vartheta_j|\mathbb{Y})\, d\vartheta_j, \tag{C.1}$$

where

$$p(\phi_g^j|\vartheta_j) = \sum_{c=1}^{\mathcal{K}_j} \frac{n_c^j}{\eta_j + \mathcal{N}_{\mathcal{A}}} \mathsf{N}(\phi_g^j|a_c^{j*}, h_{\phi c}^{2j*}) + \frac{\eta_j}{\eta_j + \mathcal{N}_{\mathcal{A}}} p(\phi_g^j). \tag{C.2}$$

Recall $p(\phi_g^j)$ is the prior predictive distribution [see (B.23)]. The posterior predictive distribution (C.1) can be approximated by

$$p(\phi_g^j|\mathbb{Y}) \approx \frac{1}{R} \sum_{r=1}^{R} p(\phi_g^j|\vartheta_j^{(r)}). \tag{C.3}$$

There is a similar expression for the posterior predictive distribution for σ_g^2.

Specific fund-observation parameters

Each fund-observation, indexed by (i,t), is associated with a regime $s_{it} = m$. Let ϕ_{it} denote the fund-regime coefficients for fund i at time t. Referring to (B.1a), we have

$$p(\phi_{it}|Y_{iT_i}, s_{it} = m) = \mathsf{N}(\phi_{it}|\overline{\mu}_{im}, \overline{B}_{im}). \tag{C.4}$$

Then,

$$p(\phi_{it}|\mathbb{Y}) = \sum_{m=1}^{M_i} p(\phi_{it}|Y_{iT_i}, s_{it} = m)\, p(s_{it} = m|\mathbb{Y}). \tag{C.5}$$

This distribution may be approximated as follows:

$$p(\phi_{it}|\mathbb{Y}) \approx \frac{1}{R} \sum_{r=1}^{R} p(\phi_{it}|Y_{iT_i}, s_{it}^{(r)}) = \frac{1}{R} \sum_{r=1}^{R} \mathsf{N}(\phi_{it}|\overline{\mu}_{im}^{(r)}, \overline{B}_{im}^{(r)}). \tag{C.6}$$

Similarly, the marginal distribution for ϕ_{it}^j may be approximated by

$$p(\phi_{it}^j|\mathbb{Y}) \approx \frac{1}{R} \sum_{r=1}^{R} p(\phi_{it}^j|Y_{iT_i}, s_{it}^{(r)}), \tag{C.7}$$

where

$$p(\phi_{it}^j|Y_{iT_i}, s_{it}) = \int p(\phi_{it}|Y_{iT_n}, s_{it})\, d\phi_{it}^{-j} = \mathsf{N}(\phi_{it}^j|\overline{\mu}_{im}^j, \overline{B}_{im}^{jj}). \tag{C.8}$$

For every fund-coefficient, indexed by (i,j), we have a sequence of posterior distributions:

$$\{p(\phi_{it}^j|\mathbb{Y})\}_{t=\tau_i}^{T_i}. \tag{C.9}$$

We will examine various features of these distributions. If there are no regime changes in going from t to $t+1$, then the two distributions will be the same. For a fund with no regime changes, the entire sequence of distributions will be the same.

Population distribution of the regime-switching probabilities

The conditional distribution (A.34a) delivers a uniform prior predictive for the probability of a regime change (regardless of the prior for k_c^*):

$$p(q_i) = \mathsf{Uniform}(0, 1). \tag{C.10}$$

Posterior

We are interested in the predictive distribution of the transition probability for the unobserved gth fund

$$p(q_g|\mathbb{Y}) = \int p(q_g|\vartheta_q)\, p(\vartheta_q|\mathbb{Y})\, d\vartheta_q, \tag{C.11}$$

where

$$p(q_g|\vartheta_q) = \sum_{c=1}^{\mathcal{K}_q} \frac{n_c^q}{\eta_q + J}\, \mathsf{Beta}(q_g|j_c^*, k_c^* - j_c^* + 1) + \frac{\eta_q}{\eta_q + J}. \tag{C.12}$$

Recall $p(\phi_g^j)$ is the prior predictive distribution [see (B.23)]. The posterior predictive distribution (C.1) can be approximated by

$$p(q_g|\mathbb{Y}) \approx \frac{1}{R} \sum_{r=1}^{R} p(q_g|\vartheta_q^{(r)}). \tag{C.13}$$

Specific regime-switching probabilities

The posterior distribution for an observed ith fund's probability of switching a regime q_i, is given by

$$p(q_i|\mathbb{Y}) = \int p(q_i|\psi_c^{q*})\, p(\psi_c^{q*}|\mathbb{Y})\, d\psi_c^{q*}, \tag{C.14}$$

where $c = z_i^q$ and $i = 1, \ldots, J$. We can approximate this posterior distribution by

$$p(q_i|\mathbb{Y}) \approx \frac{1}{R} \sum_{r=1}^{R} \mathsf{Beta}(q_i|j_{c^{(r)}}^{*(r)}, k_{c^{(r)}}^{*(r)} - j_{c^{(r)}}^{*(r)} + 1), \tag{C.15}$$

where $c^{(r)} = z_i^{q\,(r)}$.

Forecasting

Additional predictive states can be drawn via

$$p(s_{i,T_i+1}, \ldots, s_{i,T_i+H}|s_{iT_i}, q_i) = \prod_{h=1}^{H} p(s_{i,T_i+h}|s_{i,T_i+h-1}, q_i). \tag{C.16}$$

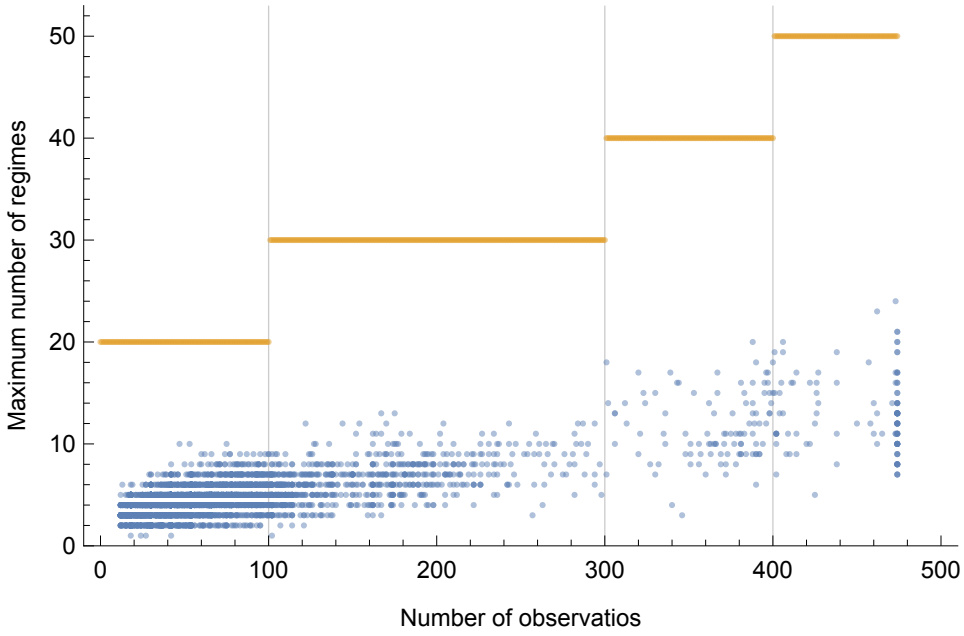


Figure 8: Maximum number of regimes versus number of observations, fund-by-fund. Maximum number allowed shown for reference.

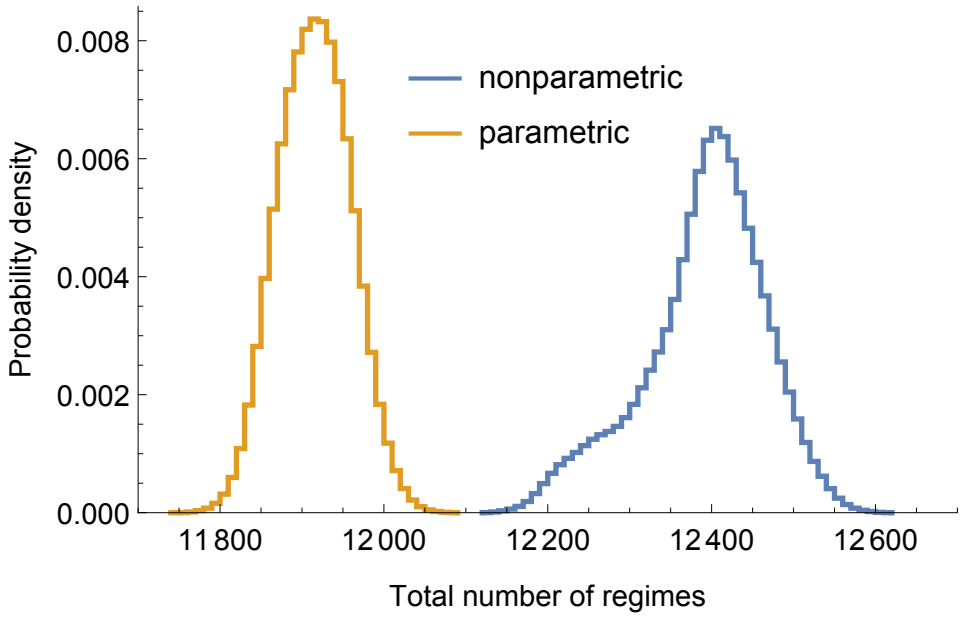


Figure 9: Smoothed histograms for total number of regimes.

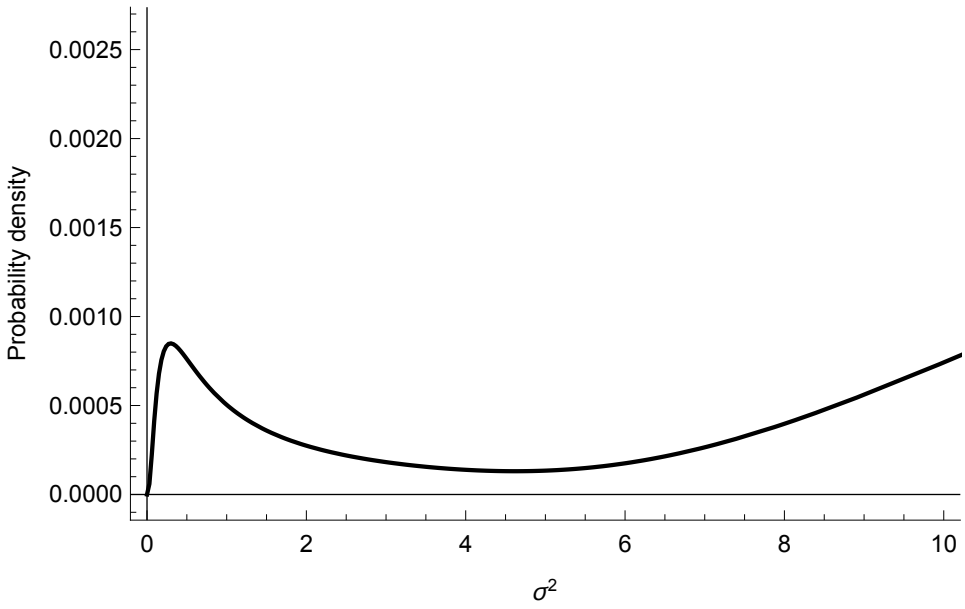


Figure 10: Detail of the posterior predictive distribution for σ^2 showing the mode located near $\sigma^2 = 1$.