Breaks in Return Predictability

Simon C. Smith\textsuperscript{a}, Allan Timmermann\textsuperscript{b}

\textsuperscript{a}USC Dornsife INET, Department of Economics, USC, 3620 South Vermont Ave., CA, 90089-0253, USA
\textsuperscript{b}University of California, San Diego, La Jolla, CA 92093-0553, USA

Draft: November 1, 2017

RESULTS ARE PRELIMINARY: PLEASE DO NOT CIRCULATE!

Abstract

We propose a new approach to forecasting stock returns in the presence of structural breaks that simultaneously affect the parameters of multiple portfolios. Exploiting information in the cross-section increases our ability to identify breaks in return prediction models and enables us to detect breaks more rapidly in real time, thereby allowing the parameters of the predictive return regression to be updated with little delay. Empirically, we find that accounting for breaks in panel return models allows us to generate out-of-sample return forecasts that are significantly more accurate than existing forecasts along statistical and economical measures of performance. Moreover, the majority of breaks in equity premiums appear to be linked to breaks in dividend growth.

Keywords: Forecasting stock returns, Panel data, Structural breaks, Bayesian analysis, Dividend growth

JEL classifications: G10, C11, C15

---

Email addresses: simonsmi@usc.edu (Simon C. Smith), atimmermann@ucsd.edu (Allan Timmermann)

The comments of Ross Valkanov and seminar participants at UCSD have been helpful. All errors are our own.
1. Introduction


Model instability may arise from a variety of sources. At the most basic level, predictability patterns in returns, if not a reflection of a time-varying risk premium, are likely to ‘self-destruct’ as investors attempt to exploit such patterns. Schwert (2003), Green et al. (2011), and McLean and Pontiff (2016) test this idea and all find evidence that abnormal returns tend to disappear after they have become public knowledge. A second possibility is that shifts in institutions, regulations, and public policy lead to shifts in the information content of the predictor variables. For example, firms may shift away from paying dividends towards repurchasing shares if taxes on dividends rise, leading to changes in the relation between dividend yields and future stock returns.

Model instability poses severe challenges to attempts at successfully predicting stock market returns. Using the full historical sample to estimate the parameters of the return forecasting model is not an attractive option if the parameters change over time as the resulting estimates may be severely biased. Conversely, using a shorter window of time (possibly after a break has occurred) leads to larger parameter estimation errors and less accurate forecasts.

Strategies of modeling the dynamics in the parameters of the return prediction model face two key challenges as pointed out by Lettau and Van Nieuwerburgh (2008). First, investors may have difficulty detecting breaks in real time. Second, and equally importantly, if a break has been detected recently, the current regime contains few observations which may lead to highly volatile parameter estimates that is likely to deliver poor forecasting performance. Overcoming these challenges has proven

\[^1\]Dangl and Halling (2012) and Johannes et al. (2014) also find improved predictability as a result of allowing for time variation in the parameters of univariate return prediction models.
difficult and Lettau and Van Nieuwerburgh (2008) find in their empirical analysis that regime shifts in the dividend-price ratio cannot be exploited to improve out-of-sample stock return forecasts. A further challenge to attempts at accounting for instability in return prediction models is that stock returns are noisy and predictors are ‘weak’ in the sense that they generate low predictive $R^2$ values. As a result, using a single time series on aggregate returns typically does not allow us to identify the timing and magnitude of any shifts to the parameter values with much precision.

This paper proposes an approach that addresses both of these concerns. We address the first challenge (slow detection of breaks) by exploiting information in the cross-section of stock returns, enabling breaks to be detected relatively quickly in real time (see Smith and Timmermann (2017)). We address the second and third challenges (imprecise model estimates) by adopting a Bayesian approach that uses economically motivated priors to shrink the parameters towards sensible values that rule out implausibly large shifts in the parameters. Specifically, following Pastor and Stambaugh (2001), we specify a prior on the intercept of the return equation which does not imply implausible Sharpe ratios. Moreover, following Wachter and Warusawitharana (2009) the prior on the slope coefficient of the predictor is centred on zero with a relatively tight variance implying that investors are sceptical about the existence of predictability. If a break has been detected recently the few data in the current regime will be unable to shift the slope estimate far from zero, but as the length of the regime increases this restriction or degree of shrinkage toward zero is reduced.

The key identifying assumption in our analysis that allow us to exploit the benefits from using panel analysis is that the timing of breaks is relatively homogenous across portfolios. To the extent that information dissemination across different segments of the market is relatively efficient, we would expect that the ability of various state variables to forecast stock returns should carry over from the aggregate stock market to individual industry portfolios. This opens up the possibility that instability in return prediction models can be more effectively detected and estimated in the context of a panel that pools return information across multiple stock portfolios. For example, if a predictor variable ceases to predict returns on the aggregate stock market portfolio, we would expect to find a similar effect on industry portfolios at approximately the same time. Exploring the simultaneous timing of breaks may allow us, both, to (i) increase our ability to detect breaks and (ii) determine their timing.

Our paper proposes a new approach to return forecasting that exploits information in the cross-section of returns to detect and adapt to instability (‘breaks’) in return
prediction models. While we assume that any breaks affect the portfolios at the same
time, we allow the intercept, slope, and variance parameters to differ across portfolios,
thus allowing for heterogeneity in the equity premium and volatility characteristics
of the individual portfolios. Exploiting cross-sectional information to estimate shifts
in model parameters turns out to be crucial to our ability to detect breaks in real
time and generate forecasts that exploit information since the most recent break.\textsuperscript{2}
Market forecasts can then be constructed as a weighted sum of \( N \) individual industry forecasts.\textsuperscript{3}

We compare our approach to several approaches that represent various degrees
of pooling of information. One approach is to separately estimate return prediction
models for individual industries and the overall market, treating each time-series as
a separate variable and ignoring possible dependencies across portfolios. This pure
time-series approach turns out to be very inefficient and does not identify breaks in
any of the time-series using the approach of Chib (1998). A second approach is to
pool the cross-sectional and time-series information in a panel, but to ignore evidence
of structural breaks. This pooling approach is also found not to generate accurate
forecasts as it ignores instability.

Our main analysis focuses on a return prediction model that uses the lagged
dividend-price ratio as a predictor variable. We jointly model predictability on 30 in-
dustry portfolios using monthly returns data over the 90-year period 1926-2015. We
also construct forecasts for the market portfolio as the value weighted average of the
industry portfolio forecasts. Empirically, we find evidence of ten breaks correspond-
ing to a little more than one break on average during each decade of our sample.
Moreover, the slope coefficient on the dividend-price ratio displays considerable vari-
ation over time with significant shifts, indicating stronger predictability over market
returns after the early seventies. Similarly, residual volatility changes markedly over
long blocks of time that get identified by our approach.\textsuperscript{4}

To help frame the question addressed in this paper, consider an investor who was
using the dividend-price ratio to predict stock returns during the financial crisis in

\textsuperscript{2}Polk et al. (2006), Hjalmarsson (2010) and Bollerslev et al. (2016) also consider predictability
of stock returns and volatility in a panel setting.

\textsuperscript{3}Other papers that have constructed aggregate forecasts by summing across components include
Ferreira and Santa-Clara (2011) and also paper by Kelly and Pruitt (2013).

\textsuperscript{4}Our approach identifies secular shifts in return volatility that in some cases lasts for up to twenty
years. Conventional approaches to model time-varying volatility tend to capture more short-lived
periods of volatility clustering. See Andersen et al. (2006) for a recent review of the literature on
volatility forecasting.
As the crisis grew deeper, investors would presumably have grown concerned about how the market instability would affect their forecasting model and whether the ability of the dividend-price ratio to predict future returns had deteriorated. As it turns out, such concerns would have been well founded. Figure 1a plots recursive estimates of the (posterior) probability that a break has occurred in the return forecasting model that uses the dividend-price ratio throughout the financial crisis, computed using our panel break model. The likelihood that a break has occurred increases smoothly from the end of 2007 to the fall of 2008 before stabilizing in early 2009. This increase in the likelihood that a break had occurred had an important effect on the slope coefficient of the dividend-price ratio (shown in Figure 1b) which declined from a level near 0.25 prior to the crisis to a level around 0.08 in early 2009. This example shows how, in real time, our approach would have detected the reduced predictability of stock returns from the dividend-price ratio and, accordingly, adjusted the sensitivity of the forecasts to this predictor variable.

The resulting return forecasts generated by our models are notably different from forecasts generated by the much smoother historical average (a benchmark proposed by Goyal and Welch (2008)) or from individual time-series models fitted to the individual portfolio returns. The latter produces much more volatile stock market returns which take on very low values—frequently below zero—during the most recent 25 years of the sample. Our panel-break forecasts generally fall in the middle between the time-series and prevailing mean forecasts.

Following earlier studies such as Campbell and Thompson (2008), Goyal and Welch (2008) and Rapach et al. (2010), we assess the predictive accuracy of our return forecasts using a variety of statistical and economic performance measures. For the market portfolio we find that the return forecasts from the panel break model are significantly more accurate than those produced by the historical average (Goyal and Welch 2008), a time-series model, or a panel model with no breaks. Specifically, we find that our panel-break approach generates significantly more accurate out-of-sample forecasts with an $R^2$ value for the market portfolio at or above 0.5 against any of the three benchmarks—a value that indicates the potential for significant economic gains using the calculations of Campbell and Thompson (2008). In an out-of-sample asset allocation analysis for an investor with mean-variance utility, we confirm that this is indeed the case. Our estimates suggest that the return forecasts from the panel break model generate certainty equivalent returns around 2% per annum relative to the benchmarks.

An important advantage from incorporating cross-sectional information from mul-
tiple portfolio return series to identify breaks is that it gives us the ability to detect breaks with a very short delay—typically just a couple of months after the break has occurred. This turns out to be crucial to explaining the gains in predictive accuracy which we document. Comparing the predictive performance of our approach to that of the benchmarks as a function of the “event time” since the most recent break, we find that our approach performs particularly well in the relative short window after a break has occurred. This suggests that one reason why our approach works so well is that it adapts more rapidly to shifts in the predictive power of individual predictor variables than conventional time-series methods.

Another benefit of our panel approach to jointly modeling returns on individual portfolios and the market is that we can evaluate not only the performance of forecasts of the market portfolio but also of the individual industries. For the 30 industry portfolios we find that our approach generates significantly more accurate forecasts between 23 and 26 cases measured relative to the three benchmarks (panel with no breaks, historical average, and time-series forecasts) without a single case in which our forecasts are significantly worse than those produced by the three benchmarks.

While our main empirical analysis focuses on a return prediction model that uses the dividend-price ratio as a predictor, we find that the strong performance of the panel-break model carries over to three other predictor variables from the finance literature, namely the one-month T-bill rate, the default spread and the term spread.

Return predictability can arise either from predictability in risk premia or from predictability in cash flow growth. While risk premia are unobservable, we can measure cash flows through dividends. We therefore undertake a separate analysis of dividend growth predictability and explore whether any breaks separately identified in the dividend process line up with the breaks found in the excess return data. We find that, indeed, the vast majority of breaks in stock returns are preceded by breaks in dividend growth. This suggests that investors’ awareness of breaks in the underlying dividend growth process is a driver of breaks in stock market returns.

The remainder of the paper is set out as follows. Section 2 lays out our panel-break approach and compares it to existing methods from the literature on return predictability. Section 3 introduces the empirical analysis and reports evidence of structural breaks. Section 4 evaluates the return forecasts of a set of industry portfolios and the market portfolio while Section 5 performs robustness checks. Section 6 concludes.
2. Methodology

This section reviews alternative approaches to capturing instability in return prediction models and introduces our novel approach which uses panel data to estimate structural breaks that simultaneously affect multiple return series. Our main specification is a heterogeneous panel model with an unknown number of breaks occurring at unknown times. While we allow the magnitude of shifts to parameters to vary across portfolios, we assume that the timing of the breaks is common in the cross-section. Our approach differs from conventional return prediction models in two regards: first, it uses panel data, as opposed to the more conventional single-equation time-series approach used throughout the literature; second, it allows for breaks.

To quantify the importance of each of these differences, we compare our approach to (i) a pure time-series approach that allows for breaks, thus highlighting the importance of using cross-sectional (panel) information; and (ii) a constant-parameter panel model that uses the same information as our approach, allowing us to gauge the importance of allowing for breaks. We explain the basic methodology below. For a more detailed exposition of the methods described in this section see Smith and Timmermann (2017), who develop the methodology applied in this paper.

2.1. Portfolio-specific Breaks and Parameters

The most general return prediction model we consider assumes that both the model parameters and breaks are unit-specific and so allows for the maximum degree of flexibility in how the individual return series are modeled. This yields a time-series model which is applied to the cross-section of the \( N \) portfolio returns on a unit-by-unit basis. Following standard practice in the return predictability literature, we focus on prediction models that include an intercept and a single predictor which can either be specific to each portfolio, \( \tilde{X}_i \), or be the same (market-wide) predictor, \( \tilde{X}_t \). We denote by \( \tilde{r}_{it+1} \) excess returns at time \( t + 1 \) on the \( i \)th portfolio and treat this as our dependent variable.

Suppose the data generating process is time-varying and subject to an unknown number of portfolio-specific structural breaks, \( K_i \), which split the sample into \( K_i + 1 \) distinct regimes for the \( i \)th portfolio. Moreover, let \( \tau_i = (\tau_{i1}, \ldots, \tau_{iK_i}) \) denote a \( K_i \)-vector of breakpoints for the \( i \)th series. The time-series model that is fitted to each
portfolio return series in the cross-section takes the form\footnote{For convenience we assume that $\tau_{i0}=0$ and $\tau_{iK_{k}+1}=T$ for all $i$.}

$$\tilde{r}_{it} = \mu_{ik_i} + \beta_{ik_i}\tilde{X}_{t-1} + \tilde{\epsilon}_{it}, \quad k_i = 1, \ldots, K_i + 1, \quad t = \tau_{ik_i-1} + 1, \ldots, \tau_{ik_i},$$

(1)

where $\mu_{ik_i}$ and $\beta_{ik_i}$ denote the intercept and slope coefficients in the $k_i$th regime and the error term is assumed to be Normally distributed $\tilde{\epsilon}_{it} \sim N(0, \sigma^2_{ik_i})$ for $k_i = 1, \ldots, K_i + 1$, and $t = \tau_{ik_i-1} + 1, \ldots, \tau_{ik_i}$.

Following existing studies such as Pástor and Stambaugh (2001), we estimate this break model using the algorithm of Chib (1998) which is perhaps the most popular Bayesian econometric breakpoint method in the economics literature. This time-series method is applied on a unit-by-unit basis to each portfolio in the cross-section.

2.2. Pooled Breaks and portfolio-specific Parameters

The model in equation (1) with both portfolio-specific parameters and break dates assumes that each cross-sectional unit is independent. However, increased power in break detection could be achieved by combining information from the cross-section of portfolio returns. The second model we consider therefore estimates breaks by pooling the information from the cross-section to identify the timing of the $K$ common breaks that separate the $K + 1$ regimes, while still estimating the parameters for each individual series

$$\tilde{r}_{it} = \mu_{ik} + \beta_{ik}\tilde{X}_{t-1} + \tilde{\epsilon}_{it}, \quad i = 1, \ldots, N, \quad t = \tau_{k-1}+1, \ldots, \tau_{k}, \quad k = 1, \ldots, K+1.$$  

(2)

Again, we assume that the error-term is Normally distributed with unit-specific variance $\tilde{\epsilon}_{it} \sim N(0, \sigma^2_{ik})$ in the (common) $k$th regime and $\tau = (\tau_1, \ldots, \tau_{K+1})$ for all $i$\footnote{The likelihood function and estimation of each model presented in this Section are detailed in Appendix A and D, respectively.}.

Some of the popular model specifications that have been considered in the literature can be obtained as special cases of equation (2). In particular, the historical average or prevailing mean model of Goyal and Welch (2008) is obtained by setting $K = 0$ and omitting $\tilde{X}_{t-1}$. Similarly, a conventional panel model with no breaks is obtained when $K = 0$. 

---

\footnote{For convenience we assume that $\tau_{i0}=0$ and $\tau_{iK_{k}+1}=T$ for all $i$.}
2.3. Correlated Effects

The 30 industry portfolio returns we are predicting exhibit high levels of correlation. Ignoring such correlations will diminish the increased break detection power obtained by using panel data rather than the individual time series of returns (Kim 2011; Baltagi et al. 2016) and so it is important to address this point.

Directly estimating the full covariance matrix of errors is infeasible given that panel break models have been shown to detect breaks with very short delay (Smith and Timmermann 2017). Our cross-sectional dimension of $N = 30$ would require estimating 525 parameters in each regime, consisting of $3N = 90$ regression parameters and $N_p = (N^2 - N)/2 = 435$ correlations. A regime duration shorter than $525/N \approx 18$ periods would therefore require estimating more parameters than we have observations within that regime. In the empirical application that follows every single break is detected with a considerably shorter delay than this.

Assume that correlations across industry portfolio excess returns are induced by a single factor (the excess return on the market portfolio, denoted $\tilde{r}_{Mrkt,t}$)

\[
\tilde{r}_{it} = \mu_{ik} + \beta_{ik}' \tilde{X}_{t-1} + \tilde{\epsilon}_{it}, \quad t = \tau_{k-1} + 1, \ldots, \tau_k, \quad k = 1, \ldots K + 1,
\]

\[
\tilde{\epsilon}_{it} = \gamma_{ik} \tilde{r}_{Mrkt,t} + \tilde{\nu}_{it}, \quad (3)
\]

in which $\gamma_{ik}$ denotes the time-varying factor loading for the $i$th portfolio in regime $k$ and $\tilde{\nu}_{it}$ denotes the idiosyncratic errors. If the factor is observed we simply add it to the regression and thereby eliminate the correlations in the errors. Even if the factor is unobserved, however, Pesaran (2006) (see also Baltagi et al. (2016)) show that cross-sectional averages of the dependent and independent variable(s) can be used as proxies for the factors. Since our predictor is aggregate we use cross-sectional averages of the dependent variable only. The weights applied across the cross-section can be specified by the user and choosing the value weights means the market portfolio can also proxy for an unobserved common factor.

The predictive regression displayed in equation (B.1) involves estimating $N(K+1)$ additional factor loadings. We detect ten breaks in our sample and this would therefore involve estimating over 300 additional parameters. Estimating so many parameters is likely to produce imprecise forecasts and so we employ an equivalent approach that recursively prefilters the data with the market portfolio excess returns available.

\footnote{The $3N = 90$ regression parameters consist of the intercept, slope coefficient and error term variance, each of which are portfolio-specific.}
at the time each forecast is made in an attempt to eliminate the correlations induced by the factor (Pesaran 2006; Baltagi et al. 2016). The details of this prefiltering approach are presented in Appendix B. A formal test conducted in Section 3.2 suggests that including the market is very successful in dealing with correlations.

2.4. Out-of-sample Return Forecasts

Using the transformed data we will obtain estimates of the intercept $\hat{\mu}_{ik}$ and slope coefficient $\hat{\beta}_{ik}$ across the $K + 1$ regimes from our predictive regression with $K$ breaks in which the number of breaks $K$ is not fixed but is estimated from the data. Avramov (2002) reports that Bayesian Model Averaging is crucial when forecasting stock returns in the presence of model uncertainty and ignoring it results in large utility losses for investors. We therefore produce our forecasts by performing Bayesian Model Averaging to account for model instability, incorporating any uncertainty surrounding both the number and timing of breaks into our portfolio forecasts.

Specifically, for the $i = 1, \ldots, N$ industry portfolio returns we generate forecasts in two stages. First, forecasts are constructed by loading the slope estimate on the raw predictive variable and adding the intercept estimate

$$\hat{r}_{i,t+1} \mid K = \hat{\mu}_{iK+1} + \hat{\beta}_{iK+1}\tilde{X}_t$$

incorporating any uncertainty surrounding the break locations while conditioning on the number of breaks $K$. Let $K_{\text{min}}$ and $K_{\text{max}}$, respectively, denote the lowest and highest number of breaks that are assigned a nonzero posterior probability by our estimation procedure. We further incorporate into the forecasts any uncertainty surrounding the number of breaks

$$\hat{r}_{i,t+1} = \sum_{K=K_{\text{min}}}^{K_{\text{max}}} p(K \mid y, X)\hat{r}_{i,t+1} \mid K.$$  

Avramov (2002) reports that Bayesian Model Averaging improves performance when forecasting stock returns in the presence of model uncertainty and investors that do not account for it will face large utility losses. Our paper performs Bayesian Model Averaging in the presence of model instability, incorporating any uncertainty surrounding both the number and timing of breaks into forecasts of portfolio returns.

---

8 For expositional ease we do not formally state the Bayesian Model Averaging that is done over the break locations.
Ferreira and Santa-Clara (2011) report that forecasting separately the three components of stock market returns - the dividend-price ratio, earnings growth and price-earnings ratio growth - can yield large improvements in predictability, while Kelly and Pruitt (2013) report that using past disaggregated value ratios can lead to improved predictability. In this spirit, we construct a forecast for the market portfolio return as the value-weighted average of the 30 industry forecasts

$$\hat{r}_{Mkt,t+1} = \sum_{i=1}^{N} w_i \hat{r}_{it+1}$$

(6)

in which \(w_t = (w_1, \ldots, w_N)\) denotes the vector of value weights on the \(N\) industry portfolios at time \(t\).

Throughout our analysis, out-of-sample return forecasts are generated recursively with an initial “warm-up” sample of ten years. Hence, the initial parameters of each model are estimated using data from July 1926 through June 1936 and a forecast is made at June 1936 for July 1936. We then expand the estimation period by one month and estimate the parameters of each model using data from July 1926 through July 1936 and make a return forecast for August 1936. This process is repeated until finally we estimate the parameters of each model using data from July 1926 through November 2015 and make the forecast for December 2015.

2.5. Prior Distributions

Our Bayesian methodology combines information in the data transmitted through the likelihood function with prior information. Details of the shape of the priors are provided in Appendix C, but we assume conventional conjugate Normal priors over the regression coefficients and inverse gamma priors on the variance parameters within each regime. The hyperparameters that determine the frequency of breaks to the coefficients are set so that a break occurs on average roughly once per decade.

It is worth emphasizing that we let the key prior parameters be economically motivated. First, given evidence on ‘no predictability’ such as Goyal and Welch (2008), we center our prior for \(\beta\) zero. Second, inspired by Wachter and Warusawitharana (2009) we explore an economically motivated prior distribution that allows investors to have different views regarding the degree to which excess returns are predictable. In the absence of breaks, if the slope coefficient \(\beta\) on the predictive variable is equal to zero, this implies no predictability, and the predictive regression is simply the ‘no predictability’ benchmark model, i.e., the historical average. A Bayesian analysis al-
lows many different degrees of predictability reflecting the scepticism of the investor as to whether excess returns are predictable. For instance, if $\beta$ is normally distributed with zero mean and variance $\sigma^2_\beta$, then setting $\sigma^2_\beta = 0$ implies a dogmatic prior belief that excess returns are not predictable, while $\sigma^2_\beta \to \infty$ specifies a diffuse prior over the value of $\beta$ implying that all degrees of predictability (and hence values of the $R^2$ from the predictive regression) are equally likely. An intermediate view suggests the investor is sceptical about predictability but does not rule it out entirely.

As noted by Wachter and Warusawitharana (2009) it is undesirable to place a prior directly on $\beta_i$ since a high variance of the predictor $\sigma^2_x$ might lower the prior on $\beta_i$ whereas a large residual variance $\sigma^2_i$ might increase it. To address this point, we first scale $\beta_i$ to account for these two variances, placing instead the prior over this ‘normalised beta’

$$\eta_i = \beta_i \frac{\sigma_X}{\sigma_i}. \quad (7)$$

The prior on $\eta_i$ is

$$p(\eta_i) \sim N(0, \sigma^2_\eta). \quad (8)$$

By (7), this is equivalent to placing the following prior on $\beta_i$

$$p(\beta_i) \sim N \left(0, \frac{\sigma^2_\eta \sigma^2_i}{\sigma_X^2} \right). \quad (9)$$

We compute $\sigma^2_X$ as the empirical variance of the predictor variable over the full sample available at the time the recursive forecast is made.\footnote{Computing $\sigma^2_X$ using only data available in the most recent regime is problematic due to the possibility of very short regimes.}

Linking the prior distribution of $\beta_i$ to $\sigma_X$ and $\sigma_i$ is an attractive feature because it implies that the distribution on $R^2$ from the predictive regression is well-defined. In population, for a single risky asset the proportion of the total variance that originates from variation in the predictable component of the return is

$$R^2_i = \frac{\beta^2_i \sigma^2_X}{\beta^2_i \sigma^2_X + \sigma^2_i} = \frac{\eta^2_i}{\eta^2_i + 1}, \quad i = 1, \ldots, N \quad (10)$$

which implies that no risky asset can have an $R^2_i$ that is ‘too large’.\footnote{Throughout our analysis we evaluate the $R^2_i$ independently of all other portfolios and thus we do not consider a prior specification that accounts for multiple risky assets.}

The informativeness of the prior is determined by $\sigma_\eta$. We refer to Wachter and Warusawitharana (2009) for a full explanation but provide the main results here for completeness. When $\sigma_\eta = 0$ the investor assigns all probability to an $R^2_i$ value of zero.
for all $i$. Figure 2 displays how investors assign more weight to a positive $R^2_i$ as $\sigma_\eta$ increases. Specifically, when $\sigma_\eta = 0.04$ the investor assigns 0.075 probability to $R^2_i$ values greater than 0.005. When $\sigma_\eta = 0.02$ the investor assigns 0.0003 probability to $R^2_i$ values greater than 0.005. When $\sigma_\eta = 0.06$ the investor assigns 0.235 probability to $R^2_i$ values greater than 0.005. For large values of $\sigma_\eta$ the investor assigns approximately equal probabilities to all values of $R^2_i$. In the main empirical analysis we consider a moderate degree of predictability by setting $\sigma_\eta = 0.04$ following Wachter and Warusawitharana (2009), but we also explore the robustness of the results when this parameter is adjusted.

It may also be desirable to specify that high Sharpe Ratios are a priori unlikely. A high absolute value of the intercept term combined with a low residual variance would imply a high Sharpe Ratio. In the spirit of P´astor and Stambaugh (1999) we multiply the prior variance of the intercept term $\sigma_\mu$, by the corresponding estimated residual variance in the $k$th regime for the $i$th portfolio $\sigma_\eta$. The intuition is as follows. The intercept term has a prior mean of zero. If the residual variance is low this reduces the overall intercept variance thereby making a large absolute intercept value and hence a high Sharpe Ratio improbable. As the residual variance increases the probability assigned to large absolute intercept values increases accordingly. We adopt a moderate prior belief in the empirical analysis by setting the prior intercept variance $\sigma_\mu$ equal to 5% following P´astor and Stambaugh (1999).

3. Empirical Results: Evidence of Breaks

This section introduces the returns data used in our study along with the predictor variables and moves on to present empirical evidence on the number of breaks identified by our approach, their location along with the resulting dynamics in the parameter estimates.

3.1. Data

As our dependent variable we use monthly returns on 30 value-weighted industry portfolios from July 1926 through December 2015 sourced from Kenneth French’s website, all computed in excess of a one-month T-bill rate. We also source monthly

\footnote{See also Avdis and Wachter (2017) who report that maximum likelihood estimation that incorporates information about dividends and prices results in an economically meaningful reduction in the equity premium estimate that is more reliable relative to the commonly used sample mean.}
returns excluding dividends from French’s website, and the 5 × 5 portfolios sorted on both size and book-to-market and size and momentum.

Our lead predictor is the dividend-price ratio, but we also consider predictors such as the one-month Treasury-bill rate, the term spread (the difference between the long term yield on government bonds and the Treasury-bill rate), and the default spread (the yield spread between BAA- and AAA-rated corporate bonds), all sourced from Amit Goyal’s website.

3.2. Testing cross-sectional dependence

Estimation is conducted on transformed data under the assumption that this transformation has eliminated any correlations across the cross-section that may be present in the raw data. We evaluate whether this is a reasonable assumption through the cross-sectional dependence test of Pesaran (2004) before prefiltering the data (see also Pesaran (2015)). This test is robust to multiple structural breaks and is therefore well suited to our framework. The test statistic follows a standard Normal distribution $CD \sim N(0, 1)$ and is computed using pairwise correlations

$$CD = \sqrt{\frac{2T}{N(N-1)}} \left( \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \hat{\rho}_{ij} \right)$$

in which $\hat{\rho}_{ij}$ is the the pairwise correlation of the residuals for series $i$ and $j$ estimated from the full sample

$$\hat{\rho}_{ij} = \frac{\sum_{t=1}^{T} e_{it} e_{jt}}{\left( \sum_{t=1}^{T} e_{it}^2 \right)^{1/2} \left( \sum_{t=1}^{T} e_{jt}^2 \right)^{1/2}}$$

and $e_{it}$ is the residual from the OLS time series regression for the $i$th series

$$e_{it} = \tilde{r}_{it} - \tilde{\mu}_i - \tilde{\beta}'_i \tilde{X}_{t-1}.$$ 

The CD statistic for our lead predictive variable, the dividend-price ratio, is equal to 168.84 and we therefore conclusively reject at the 1% level the null of zero cross-sectional dependencies in the raw data. Performing the same calculations using the prefiltered data $\tilde{y}$ and $\tilde{X}$ we obtain a CD test statistic of 2.39 which means the null of zero cross-sectional dependencies can no longer be rejected at the 1% level. Furthermore, prefiltering the data has reduced the average of the absolute pairwise
correlations from 0.74 to 0.13. The prefiltering approach is therefore very successful in removing the strong cross-sectional dependencies in the data suggesting that any cross-sectional dependence that remains is likely to be weak. In a large panel setting \((N > 10)\) like the one we have here, Pesaran (2015) notes that only strong cross-sectional dependence can compromise inference. For example, in portfolio analysis, for full diversification of idiosyncratic errors we only require weak cross-sectional dependence, not independence. Any cross-sectional dependencies that may remain after prefiltering are likely to only be weak and therefore are not a concern for our subsequent empirical analysis.

3.3. Evidence of breaks

We first consider the evidence of breaks in the return prediction model as identified by our approach. To this end, the top panel in Figure 3 plots the posterior probability distribution for the number of breaks estimated on the full sample of 90 years of data for the model that uses the lagged dividend-price ratio as a predictor. The mode (and mean) for the number of breaks is 10, with approximately 90% of the probability mass distributed between 9 and 10 breaks. These estimates suggest a break occurring roughly every nine years.

The lower panel in Figure 3 plots the posterior probability for the location of the breaks. The timing for most of the breaks appears to be quite well defined with clear spikes in the posterior probabilities in 1929, 1933, 1972, 1998, and 2008. Thus, the break dates coincide with major economic events such as the Great Depression, the oil price shocks of the 1970s, and the financial crisis of 2008. Interestingly, the posterior probability mass is quite disperse during the recent financial crisis, indicating that its effect on different industry portfolios was not confined to a single month but diffused gradually through time. Note also that there are long periods without any evidence of model instability, e.g., the twenty year period from 1950 to 1970.

The breakpoints identified by our panel approach are very different from the breakpoint estimates obtained from the breakpoint algorithm of Chib (1998) applied to the univariate time series of returns on the individual industry portfolios. In fact, for each of the industry portfolios the univariate breakpoint model fails to detect a single break, always favoring the model with zero breaks placing, on average, 91.21% of the posterior model probability on the no break model. This suggests that the

---

\(^{12}\)The results (not shown) are similar for the other three predictors.
tests have too weak power to identify breaks off individual return series when based on information on the evolution in returns alone.\textsuperscript{13}

### 3.4. Evolution in Parameter Estimates

Having identified the location of the breakpoints, we next turn to the dynamics in the estimates of the model parameters. To this end, Figure 4 graphs the evolution in the slope coefficient and error-term standard deviation of the market portfolio over the out-of-sample period. The estimates at each time are computed as the value-weighted average of the recursively estimated parameters on the 30 industry portfolios. While always positive, the estimated slope coefficient on the dividend-price ratio (top window) changes considerably over the sample, taking values up to 0.4 in the late forties. The plot shows that although we allow the estimated coefficients to jump from one regime to another, and sometimes observe rapid shifts in the parameters, there are also notable episodes with gradual shifts in the slope coefficients such as from 1969-1971. These gradual shifts arise during times with greater uncertainty about the occurrence of a break.

The volatility parameter (shown in the lower window) shows notable peaks between the great depression and the end of World War II and after the financial crisis. Conversely, the market volatility parameter is notably lower for a long spell between 1947 and 1970.

Figure 4 also plots parameter estimates for the panel model with no breaks and the univariate time-series model fitted to market returns alone. The volatility estimates from the panel model with no breaks and time series model with breaks are very smooth hovering around 0.09 through the sample. The estimated slope coefficient from the panel model without breaks is generally smaller than the values obtained from the panel model that allows for breaks. The variation in the estimates generated by the univariate time-series model fitted to market returns are not a result of breaks, since no breaks are identified, but, rather, a result of the recursive updates to the parameter estimates.

We conclude the following from these plots. First, allowing for breaks and including information on multiple portfolios in a panel setting appears to make a considerable difference to the estimates which behave very differently from estimates

\textsuperscript{13}Pastor and Stambaugh (2001) identify breaks in returns based on assumptions about joint movements in the mean and variance of returns.
based either on a univariate break model fitted on the individual industry portfolio returns or from estimates fitted to a panel model without breaks. Second, the estimated breaks for the return model are sometimes driven by the volatility parameter, at other times get identified from the relationship between return movements and movements in the lagged dividend-price ratio, i.e., the slope coefficient.

Ultimately, we are interested in how shifts in parameter estimates affect the return forecasts. This cannot be gleaned from the plots of the estimated coefficients in Figure 4 because some of the shifts in the (mean) coefficients may be partially offsetting and both the coefficient estimate and the dividend-price ratio vary at the same time.

To show how changes to the estimated parameters affect the return forecasts, we therefore study the return forecasts, in each case generated recursively or ‘out-of-sample’. We first consider return forecasts of the market portfolio in more detail. The out-of-sample return forecasts from the heterogeneous panel model with (dashed red line) and without (solid purple line) breaks and the prevailing mean model (dotted black line) are shown in Figure 5. The forecasts generated from the prevailing mean model are much smoother than the other ones. Return forecasts from the two panel models display higher volatility than the prevailing mean model. Still, the forecasts from the two panel models are quite different, indicating the importance of allowing for breaks.

3.5. Real-time detection of breaks

A key challenge when generating return forecasts in a setting that accounts for breaks is how quickly the model is able to identify breaks in real time. Severe delays in breakpoint detection is likely to lead to poor forecasting performance, particularly if the distance between breaks is relatively short, causing some regimes to be overlooked altogether. Conversely, if shifts to parameter values can be identified with little delay, this opens the possibility of improved forecasting performance. The ability to detect breaks in real time is of central importance to investors who must re-allocate their portfolios in a timely manner.

To shed light on this issue, Figure 6 plots the break dates as they are estimated in real time. The real-time breakpoint detection performance of the model with pooled breaks and portfolio-specific parameters works as follows. The initial model is estimated using the first ten years of data. Subsequently, the estimation window is expanded by one month and the model is re-estimated until we reach the end of the sample, recording the break dates at each time. The vertical line in the figure marks
the first period at which the model is estimated given the initial training window of ten years (120 monthly observations) while the 45 degree line (to the right of the vertical line) marks the points at which a break could first be detected, corresponding to a delay of zero. Empty circles on the graph mark the break dates as estimated in real time with horizontal bands of circles indicating that an initial break date estimate is confirmed to have occurred as subsequent data arrive. The figure is dominated by these bands whose initial points start with only a short delay from the 45 degree line, clearly displaying the ability of the procedure to rapidly detect the onset of a break. Conversely, when initial break estimates are not supported by subsequent data points, as indicated by isolated circles outside the horizontal bands, this is indicative of “false alarms”. There are not too many instances in which the approach detects what subsequently turns out to be spurious breaks.

Lettau and Van Nieuwerburgh (2008) and Viceira (1997) find evidence of instability in time-series predictive regressions of the aggregate market return on the dividend-price ratio. They find, however, that such instability cannot be exploited out-of-sample because their univariate method is unable to detect breaks in real time. Figure 6 shows that, by incorporating cross-sectional information from returns on multiple portfolios, our panel break procedure has increased break detection power relative to the time-series approach.

Short delays in detecting breaks to the parameters of the return prediction model are, thus important to a successful forecasting strategy. To further highlight this point, Figure 7 plots the number of months before a break was detected in real time, measured relative to the full-sample (ex-post) estimate of the break date. The majority of breaks in the dividend-price ratio model were detected within five to eight months of their occurrence, with the longest delay being 9 months.

The ability of our panel breakpoint approach to identify breaks with relatively little delay helps explain its good forecasting performance. Moreover, it stands in marked contrast to the long delays typically associated with breakpoint modeling in the context of univariate time-series.

4. Evaluation of Return Forecasts

This section compares the predictive performance of our heterogeneous panel break model with a univariate time-series break model, a heterogeneous panel model without breaks, and the simple historical average, the latter serving as a ‘no predictability’ benchmark. We report both statistical and economic measures of forecasting perfor-
mance, the latter based on how a risk averse mean-variance investor would utilize the forecasts from the different return prediction models.

4.1. Measures of Predictive Accuracy

We evaluate the forecasting ability of each of the models relative to each of the benchmark models for the \(i\)th portfolio through the commonly used out-of-sample \(R^2\) measure proposed by Campbell and Thompson (2008):

\[
R^2_i = 1 - \frac{MSE_{i,Pbrk}}{MSE_{i,Bmk}}.
\]

Here \(MSE_{i,Pbrk}\) denotes the mean squared forecast error for the \(i\)th portfolio obtained from the panel break model and \(MSE_{i,Bmk}\) denotes the mean squared forecast error for the \(i\)th portfolio obtained from the benchmark model in question. A positive \(R^2\) value indicates outperformance relative to the benchmark model, while a negative value indicates underperformance.

To evaluate whether the difference in the predictive accuracy of two sets of forecasts is statistically significant, we use two measures namely the mean squared error (MSE) differential proposed by Diebold and Mariano (1995) and the MSE-adjusted test statistic of Clark and West (2007). The Diebold-Mariano test is obtained by regressing the squared forecast error differentials of the panel break model relative to those produced by a given benchmark on an intercept and computing the resulting \(t\)-statistic.

To implement the approach of Clark and West (2007) for the \(i\)th portfolio, first define

\[
f_{it+1} = (\hat{r}_{it+1} - \hat{r}_{Bmk,it+1})^2 - [(\hat{r}_{it+1} - \hat{r}_{Pbrk,it+1})^2 - (\hat{r}_{Bmk,it+1} - \hat{r}_{Pbrk,it+1})^2] \tag{15}
\]

in which \(\hat{r}_{Bmk,it+1}\) denotes the forecast of return for the \(i\)th portfolio at time \(t + 1\), generated at time \(t\) from the benchmark model which is either the prevailing mean, the time series break model or the heterogeneous panel no-break model, \(\hat{r}_{Pbrk,it+1}\) denotes the predicted return for the \(i\)th portfolio at time \(t + 1\) from the panel break model generated at time \(t\), and \(\tilde{r}_{it+1}\) denotes the realised return for the \(i\)th portfolio at time \(t + 1\). Letting \(m = 120\) denote the initial training period, a \(p\)-value is obtained with the standard normal distribution by regressing \(f_{im+1}, \ldots, f_{iT}\) on a constant and computing the corresponding \(t\)-statistic.
4.2. Out-of-sample return forecasts

To evaluate the accuracy of the return forecasts, Figure 8 plots the cumulative sum of squared error differences (CSSED) produced from our panel method forecasts versus those generated by each of the benchmark models

\[
CSSED_{it} = \sum_{\tau=1}^{\tau=t} (e_{Bmk,i\tau}^2 - e_{Pbrk,i\tau}^2),
\]

in which \(e_{Bmk,i\tau}\) and \(e_{Pbrk,i\tau}\) denote the respective forecast errors from the benchmark model in question and our panel break model for the \(i\)th portfolio at time \(\tau\). Positive and rising values of the CSSED measure represent periods where the panel break model outperforms the respective benchmarks, while negative and declining values suggest that the panel break model is underperforming. Moreover, if the performance of the panel break model measured against the benchmark is dominated by a few observations, this will show up in the form of sudden spikes in these graphs. In contrast, a smooth, upwardsloping graph indicates more stable outperformance of the panel break model measured against the benchmark.

Figure 8 presents plots of the CSSED values for the market portfolio and three representative industries (oil, financials and telecommunications). The plots show that the heterogeneous panel model with breaks consistently outperforms its competitors over the 80-year sample. For the market portfolio (top left hand corner), the CSSED curve for the panel model with breaks measured relative to the prevailing mean model rises throughout the out-of-sample period with no long spells of underperformance. The strong performance against the historical average is particularly impressive given that this benchmark has been found by Goyal and Welch (2008) to be very difficult to beat out-of-sample. A similarly strong performance is seen for the panel breaks model measured against either the panel model without breaks or against the univariate market model that allows for breaks.

Similar improvements in predictive accuracy from the panel breaks model are seen in the plots for the three industry portfolios displayed in Figure 8. The plots continue to show clear and consistent improvements against the prevailing mean and univariate time-series model while the improvements against the no-break panel model are more concentrated towards the last 15 years of the sample for the oil and telecommunications industries.

Figure 9 plots histograms of the \(R_{OoS}^2\) values for each of the thirty industry portfolios and the market portfolio based on comparisons of the forecasting performance
of our proposed panel breaks model relative to the three benchmark models. For the 31 portfolios our method outperforms all three benchmarks 29 times. Moreover, many of the $R^2_{OoS}$ values are economically large: Campbell and Thompson (2008) estimate that even an $R^2_{OoS}$ value as small as one-half of one percent on monthly data is economically large for a mean-variance investor with moderate risk aversion.

Table 1 uses the test statistic of Diebold and Mariano (1995) to evaluate the statistical significance of the relative performance of the panel break model against the three benchmarks. The table shows that the outperformance associated with the panel break model is significant at the 10% level for 25, 24, and 26 of the 30 industry portfolios and the market index compared to the predictive performance of the heterogeneous panel model with no breaks, the prevailing mean, and the time-series break model. Using the procedure of Clark and West (2007) this outperformance is significant at the 10% level for 27, 26, and 27 of the 31 portfolios relative to the no-break panel model, the prevailing mean model and the univariate time-series model, respectively. Conversely, the panel break model does not underperform relative to these benchmarks at the 10% level for any of the 31 portfolios.

These findings underline that the improvements in predictive accuracy that we observe for the panel breaks model is not simply a result of expanding the information set from a univariate time-series setting to a panel setup that incorporates cross-sectional information. Conversely, allowing for breaks in a univariate setting also does not produce nearly the same gains in predictive accuracy as the panel breaks model. Rather, it is the joint effect of using cross-sectional information in a panel setting and allowing the return forecasts to account for breaks that generates improvements in predictive accuracy.

Moreover, the results suggest that our panel model with breaks has the ability to adapt to breaks, and thus handle model instability, while simultaneously reducing the effect of estimation error which has so far plagued real-time (out-of-sample) return forecasts, see Lettau and Van Nieuwerburgh (2008).

4.3. Forecasting Performance in the Aftermath of Breaks

To the extent that pooling cross-sectional information helps the panel-break model speed up learning, we would expect forecasting performance to be particularly good in the immediate aftermath of a break, particularly if the break is large in magnitude.

Figure 10 graphs the cumulative difference in the sum of squared errors as a function of the time since the initial break detection, measured in months, i.e., in
break point ‘event time’. Specifically we compute the squared forecast errors in each month following the detection of each break in the out-of-sample period and then take the mean of the squared errors in each period across the breaks, i.e. in the first month following break detection we average across the squared forecast errors in the period immediately following each of the breaks that are detected over the out-of-sample forecasting window. Across all four portfolios, our panel-break method outperforms the competing benchmarks by most in the short period after a break is detected, demonstrating the usefulness of using our panel procedure to detect the onset of a break more quickly in real time.

4.3.1. Economic Utility from Return Forecasts

In addition to evaluating the statistical significance of improvements in the predictive accuracy of forecasts from our panel break forecasts relative to a variety of benchmarks, it is important to evaluate their economic significance. For each of the 31 portfolios (including the market) we therefore compute the utility gain to a mean-variance investor who at each period allocates his portfolio between one or possibly multiple risky asset(s) and risk-free T-bills.\(^{14}\) In this one-period-ahead forecasting exercise, at time \(t\) the mean-variance investor allocates a portion of his portfolio to equities in period \(t + 1\)\(^{15}\)

\[
\begin{align*}
  w_{Bmk,t} = \frac{1}{A} \frac{\hat{r}_{Bmk,t+1}}{\hat{\sigma}^2_{\text{stk},t+1}},
\end{align*}
\]

where \(\hat{\sigma}^2_{\text{stk},t+1}\) is the stock variance estimate for time \(t + 1\) using information available at time \(t\). Following Campbell and Thompson (2008) we use a five year rolling window of monthly stock returns to estimate the variance of stock returns, and assume a risk aversion coefficient of \(A = 3\). From investing with the benchmark model in question the investor realises an average utility of

\[
\hat{v}_{Bmk} = \hat{\mu}_{Bmk} - \frac{A\hat{\sigma}^2_{Bmk}}{2},
\]

where \(\hat{\mu}_{Bmk}\) and \(\hat{\sigma}^2_{Bmk}\) denote the sample mean and variance of the portfolio formed over the out-of-sample period using the benchmark model in question.

Similarly, we compute the average utility derived by the same investor when fore-
casting using our panel breaks model. Under this scenario, the investor will assign the following weight to the stock market portfolio

\[
   w_{\text{Pbrk},t} = \frac{1}{\hat{A} \hat{\sigma}^2_{\text{Stk},t+1}}. \tag{19}
\]

The investor realises an average utility of

\[
   \hat{v}_{\text{Pbrk}} = \hat{\mu}_{\text{Pbrk}} - \frac{\hat{\sigma}^2_{\text{Pbrk}}}{2}, \tag{20}
\]

where \( \hat{\mu}_{\text{Pbrk}} \) and \( \hat{\sigma}^2_{\text{Pbrk}} \) denote the sample mean and variance of the portfolio formed over the out-of-sample period using the forecasts from our panel breaks model.

The utility gained from forecasting with our panel break model relative to the benchmark model is obtained as the difference between equations (20) and (18). Multiplying this number by 1200, we obtain the certainty equivalent return (CER) expressed in annualised percentage terms.

4.3.2. Empirical results

Figure 11 shows the distribution of utility gains across the 31 portfolios that we consider here. Allocations are based on the recursively generated out-of-sample return forecasts that use the dividend-price ratio as a predictor. Again we compute CER values for the panel breaks model relative to the utility obtained from the three benchmark specifications, i.e., a panel model with no breaks, a prevailing mean model computed for each portfolio, and separate time-series forecasts fitted separately to the individual portfolios, in each case allowing for breaks.

The plots show that the panel model with breaks generates positive CER values for at least 28 of the 31 portfolios, regardless of the benchmark. Moreover, the estimated utility gain from using the panel-break forecasts is generally economically large. For the market portfolio it is 2% per annum when measured against the forecasts from the no-break panel model or the time series model with breaks and it exceeds 1.10% relative to the prevailing mean model.

If breaks in the model parameters do not strongly affect a particular industry portfolio, it is unlikely that a model that accounts for such breaks can significantly outperform a model that ignores breaks. To see if this holds, Table 2 explores the relation between the magnitude of the break, as measured by the mean squared difference between the forecasts from the panel models with and without breaks, and the utility gains for that portfolio, again measured using the panel models with
and without breaks. The calculations again assume a mean-variance investor with a
coefficient of risk aversion of three. To keep it simple, we show only results for the
upper and lower quartile of industries, ranked by mean squared forecast difference.

We find that those industries for which breaks have the biggest effect on the
forecasts (upper quartile) generally lead to higher utility gains both in absolute and
relative terms, while industries whose return forecasts are least affected by breaks
(lower quartile) are associated with the smallest utility gains.

4.3.3. Industry Allocation Analysis

We also explore the utility gain of a mean-variance investor who at each period
allocates his wealth between the risk-free rate and a risky portfolio constructed from
the 30 industry portfolios (see e.g. Avramov and Wermers (2006) and Banegas et al.
(2013)). Let \( \tilde{r}_{p,t+1} \) denote the return on the risky portfolio at time \( t+1 \) in excess of the
risk-free rate \( r_{f,t+1} \). The return on the risky portfolio is constructed from the returns
on the 30 industry portfolios at time \( t+1 \), \( \tilde{r}_{t+1} \), and the corresponding portfolio value
weights \( w_{t+1} \). Using our panel break model at each time \( t \) we determine the weight
vector \( \omega \) to allocate among the 30 industry portfolios in the next period. Numerical
methods are used to compute \( \omega \) that maximizes the expected utility function

\[
E[U(\tilde{r}_{p,t+1} | A)] = r_{f,t} + \omega' \tilde{r}_{t+1} - \frac{A}{2} \omega' \tilde{S}_t \omega,
\]

subject to the summability constraint \( \sum_{i=1}^{N} \omega_i = 1 \), and \( \omega_i \in [0,1] \) for \( i = 1, \ldots, N \) to
preclude any leverage or short selling of each individual portfolio.\(^{16}\) The covariance
matrix at each time \( t \), \( \tilde{S}_t \), is estimated using the residuals from the return prediction
model up to time \( t \).

The vector of portfolio weights chosen to maximize expected utility at time \( t \),
\( \omega \), is plugged back into (21) to obtain the realized utility for time \( t \). This process
is repeated for each time period out-of-sample for each forecasting model we con-
der.

The difference between the utility derived from our panel break model and
the benchmark model is multiplied by 1200 to express it as an annualised percentage
CER.

Table 3 shows the average industry allocations under the four different return

\(^{16}\)Imposing constraints on the portfolio weights is akin to applying shrinkage on the variance-covariance estimates which can lead to performance improvements in mean-variance analysis. See Jagannathan and Ma (2003) and DeMiguel et al. (2007).
forecasts, in each case averaged through time.\textsuperscript{17} There are some notable differences across the different forecasts. For example, the average allocation to services is only 16\% under the historical average, but close to 40\% under the three other approaches. Conversely, the two panel models allocate substantially more (13\% and 24\%, for the break and no-break models, respectively) to the smoke industry than the historical average (9\%). In turn, the time-series model allocates far more (23\%) to telecommunications than the other models, with the allocation from the panel-breaks model (10\%) in a distant second place.

The top panel of Table 4 reports the resulting utility gains from these optimized allocations across industry portfolios. Specifically, the table shows the annualized CER values for the panel break model measured relative to the three benchmarks, in each case using the dividend-price ratio as a predictor (top line). Relative to the historical average forecasts, the CER value of the panel break model is 2.19\% per annum. Moreover, the CER value of the panel break model remains large–approximately 2\% per annum–when measured against the univariate time-series and no-break panel model.

Furthermore the improved predictive performance in the immediate aftermath of a break being detected translates into even larger utility gains during these periods. Table 4 also reports utility gains computed using only those time periods that occur within two years of a break first being detected (‘After breaks’). The annualized CER value of 3.02\% is even higher reflecting the ability of our approach to exploit the rapid detection of breaks for utility gains.

These values suggest that the panel-break forecasts of returns on the individual industry portfolios could have been used to generate economically meaningful improvements over forecasts from the three benchmarks.

5. Sources of breaks in return predictability

Return predictability can arise from two principal sources, namely predictability of cash flow growth or predictability of equity risk premia. While equity risk premia are not directly observable, we can obtain good proxies for cash flows. This section therefore explores the hypothesis that the breaks identified in the return prediction model are driven by shifts in the underlying dividend growth process.

Predictability in dividend growth is still a hotly disputed area, in part because

\textsuperscript{17}To preserve space, we omit industries that have an allocation of less than 0.01 for every model.
interpretation of variation in the dividend-price ratio depends on this issue. On one side, Cochrane (2008) argues that there is little evidence that dividend growth can be predicted. Conversely, Binsbergen et al. (2010) report annual out-of-sample $R^2$ values of 13.9-31.6% for dividend growth rates when using a present value model. See also Chen (2009) and Kelly and Pruitt (2013) for evidence of dividend growth predictability.

One explanation of disagreements regarding dividend growth predictability could be that it varies over time, i.e., dividends are highly predictable in certain regimes, while largely unpredictable in other ones.\textsuperscript{18}

To address this question, we first extract a monthly dividend series for each industry portfolio from which we compute the year-on-year dividend growth at time $t$ as the difference in the log cumulative monthly dividend from time $t - 11$ through $t$ and the log cumulative monthly dividend from time $t - 23$ through $t - 12$. We run a predictive regression with the dividend growth series for the 30 industry portfolios as the dependent variable and an intercept, autoregressive term and a lagged predictor on the right-hand-side.

Table 5 displays the posterior mean and standard deviation of the estimated intercept, AR(1) slope, dividend-price ratio slope, and the volatility obtained from our heterogeneous panel break model. The predicted dividend growth rate varies in a wide range that span high-growth states such as that in regime eight which has a large positive intercept and AR(1) coefficient and states with negative expected dividend growth in the form of states one and three which coincide with the Great Depression. The AR(1) coefficient is highly significant and positive in nine out of ten regimes. Similarly, the estimated slope of the dividend-price ratio is negative in nine of ten regime as we would expect if forecasts of higher future dividend growth lead to higher current prices and thus a smaller dividend-price ratio.\textsuperscript{19}

The bottom row of Figure 12 displays the estimated break dates identified by our panel model with breaks using the dividend-price ratio as the predictor. Blue diamonds mark the posterior modes of the break dates estimated from the heterogeneous

\textsuperscript{18}Our analysis is also relevant for past work on the predictive power of the dividend-price ratio (or dividend yield) over stock returns. Lettau and Ludvigson (2005) find that forecasts of dividends and forecasts of excess stock returns covary over the business cycle implying that positively correlated fluctuations in expectations of both dividend growth and returns have counterbalancing effects on the log dividend-price ratio.

\textsuperscript{19}The only regime where the AR(1) and dividend-price ratio coefficients have the wrong sign is regime 2. This reversal of sign can sometimes happen in short-lived regimes due to collinearity between the regressors.
panel break model fitted to dividend growth, while red diamonds mark the modes of break dates fitted to the model for excess returns. First, note that the break in 1929 is now undetected. This is likely because our sample now begins in 1928 and it can be difficult to detect breaks at the very beginning of the sample (Bai and Perron 1998). The remainder of the posterior modes of estimated break dates are very close to their original modes when excess returns were the dependent variable. In fact, every remaining break from the return model is estimated within one year of the original break date estimate (shown in brackets) except for an additional break identified in 1986 and the break in 2010 being overlooked. The results across the other three predictive variables listed in the upper rows are even stronger. The similarity among the break dates suggests that the instabilities in the return prediction model that are crucial to producing better forecasts of excess returns arise from permanent shifts in the dividend growth process.

Interestingly, the breaks in the dividend growth regression lead the breaks in the excess return regression. Ignoring the first break identified by the excess return regression (since the corresponding break is not detected by the dividend growth regression) the break in the dividend growth leads the break in returns by an average of 23 months. If we further ignore the first of the two breaks in the mid-1990s (for which the break dates from the two models are very different) the dividend growth model detects breaks that lead the return model by 12 months, on average. This evidence is suggestive that prior breaks in the dividend growth rate do, at least in part, explain the observed breaks in the return prediction models.

6. Robustness of results

This section undertakes a number of exercises to (i) investigate whether breaks are common across portfolios; (ii) establish the general validity of our empirical findings to other predictor variables from the finance literature; and (iii) explore the sensitivity of our findings to changes in the model specification such as allowing the breaks to affect some, but not all, parameters or imposing homogeneity in the regression coefficients across the portfolio returns.

6.1. Are breaks common across portfolios?

We assume breaks are common throughout. To investigate whether this assumption is reasonable we run our estimation procedure using the same model but replacing the
excess returns on the 30 industry portfolios with a range of different excess portfolio returns. We begin with the $5 \times 5$ portfolios sorted on size and book-to-market\textsuperscript{20}

The middle set of points in Figure 13 displays the results when using the dividend-price ratio as the predictor. The posterior modes of all but one of the breaks (the 1987 break which was previously identified in 1996) are estimated within one year of the original posterior modes, while the break in 2010 is overlooked altogether. The similar break dates imply that the our common break assumption is reasonable. We further run the same procedure using $5 \times 5$ portfolios sorted on size and momentum. The highest set of points in Figure 13 displays the results. One break is detected in 1933 while previously two breaks were detected in 1932 and 1934, respectively. Otherwise the break dates are very similar. For both portfolio sorts, the results are similar across the other three predictors we consider. Given the slight differences in break date estimates there is some evidence that information takes a relatively short amount of time to diffuse through portfolios, implying that markets are relatively efficient and ultimately are common break assumption is a reasonable approximation of reality.

6.2. Results using other predictors

Up to this point we have focused our analysis on a return prediction model that uses the dividend-price ratio as a predictor variable. However, a variety of other predictor variables have been proposed in the literature so we next consider three additional predictor variables in common use in the literature on predictability of stock market returns, namely the one-month T-bill rate, the default spread, and the term spread. Robustness of the performance of our panel breaks approach for these additional predictors will increase our confidence in the broader success and applicability of our approach.

To this end, we undertake the same analysis as that conducted for the model that uses the dividend-price ratio as a predictor. Again, all forecasts are generated recursively out-of-sample with a ten-year warm-up period.

To preserve space we do not show as many details of the analysis as we did for the model that uses the dividend-price ratio as a predictor. However, it is worth noting that a similar number of breaks is identified for the models that use the three other predictor variables.

\textsuperscript{20}Since the power to detect breaks increases with the size of the cross-section $N$ it is preferable to select portfolio sorts such that $N$ is close to the original value of 30. Portfolio sorts involving investment or profitability begin only in 1963 and thus are not suitable for our robustness check.
predictor variables. For example, the model that uses the T-bill rate as a predictor identifies 9 breaks, the locations of which are very similar to those for the model based on the dividend-price ratio (Figure 14a). Moreover, the ability to detect these breaks with little delay continues to hold for this predictor (Figure 14b).

In fact, the ability of our approach to identify breaks in ‘real time’ holds across all predictors. To illustrate this, Figure 7 shows the distribution of the delays in detecting breaks for all four predictors. The vast majority of breaks are detected with a delay of a few months although a few breaks get detected with a longer delay.

Figure 15 summarizes our findings through plots of the difference in the cumulative sum of squared errors for our panel break approach measured relative to that of the three benchmarks described earlier. For simplicity we focus our analysis on the market portfolio, but similar plots are obtained for the majority of industry portfolios. We see clear evidence that our panel break approach consistently produces more accurate forecasts than the alternatives.

Again we supplement the graphical analysis with more formal results. Table 1 shows that the panel breakpoint model continues to perform very well for the three alternative predictors, generating Diebold-Mariano test statistics that indicate significant improvements over the three alternative benchmark specifications for between 20 and 25 of the 31 industry and market portfolios and significant underperformance for only one case out of a total of 372 pair-wise forecast comparisons across predictor variables and benchmarks. Even stronger results are obtained for the Clark-West test statistic for which the panel-break model outperforms the three benchmarks for between 22 and 28 of the 31 portfolios.

Turning to the economic value associated with the panel break forecasts, the estimated CER values in Table 4 give evidence that the forecasts from the panel break model, when implemented in a simple mean-variance investment strategy, continue to generate utility gains in the neighborhood of 2% per annum measured relative to an investment strategy based on the historical average, a pure time-series (break) model, or a no-break panel model.

These results corroborate the more detailed analysis of the model that used the dividend-price ratio as a predictor of stock returns and so suggest that our findings are not sensitive to using a particular predictor.
6.3. Sensitivity of results to priors

Table 6 displays the results of a prior sensitivity analysis. Specifically, we adjust one hyperparameter at a time from our baseline specification and re-estimate the model. First, adjusting the hyperparameter $c$ that controls our prior expected regime duration from 20 years to 10 (40) years results in the detection of 12 (8) breaks. The significant outperformance of each of the three benchmarks across the majority of the portfolios is unaffected by adjusting this prior; neither it is affected by reasonable adjustments to the hyperparameter $b$ that controls the prior volatility or $\sigma_{mu}$ that controls the prior dispersion of the intercept. Adjusting the hyperparameter $\sigma_{\eta}^2$ to economically plausible values also has little impact on the forecasts. However, allowing this parameter to become very large which implies any degree of predictability and hence out-of-sample R-squared value is equally likely leads to the breakdown of the forecasts corroborating the findings of (Wachter and Warusawitharana 2009).

7. Conclusion

A large literature on return predictability has found evidence of model instability, i.e., the phenomenon that the parameters of the return prediction model change over time. Such model instability helps explain why out-of-sample return forecasts often are found to perform poorly compared to full-sample estimates of return predictability. While this model instability is, thus, known to affect return forecasts, exploiting such breaks has so far proved elusive due to the noisy nature of returns and the low predictive power of return prediction models.

In this paper we developed an approach that exploits cross-sectional information on multiple portfolios to detect breaks from the joint dynamics of the return series. While our approach assumes the timing of the breaks to be the same, the mean and variance parameters are allowed to differ across the individual portfolio return series. Empirically, we find that using cross-sectional information on portfolio returns in a panel model that allows for breaks substantially increases our ability to detect breaks in return prediction models. Most importantly, our approach has the ability to accurately detect breaks in real time with very little delay. This means that out-of-sample return forecasts generated by the panel model with breaks are consistently more accurate than forecasts generated by a variety of extant approaches from the literature. Interestingly, gains in predictive accuracy over such benchmarks are particularly large shortly after a break has occurred, suggesting that the model may pick up a “learning
premium” due to its ability to adapt to parameter shifts more rapidly than existing approaches.
References


Appendix A. Likelihood function

The intercept $\mu$, slope coefficient $\beta$, and error-term variance $\sigma^2$ undergo permanent shifts following each break. Let $l_{ik} = \tau_{ik} - \tau_{ik-1}$ denote the duration of the $k$th regime which contains the observations $(\tau_{ik-1+1}, \ldots, \tau_{ik})$, and let $l = (l_{i1}, \ldots, l_{iK_i+1})$. Let $\mu_i = (\mu_{i1}, \ldots, \mu_{iK_i+1})$, $\beta_i = (\beta_{i1}, \ldots, \beta_{iK_i+1})$, $\sigma^2 = (\sigma^2_1, \ldots, \sigma^2_{K_i+1})$, $\theta_i = (\mu_i, \beta_i)$, and $\Theta_i = (\mu_i, \beta_i, \sigma^2_i)$. Assuming the errors follow a Gaussian distribution with zero mean and regime-specific variance $\sigma^2_{ik}$, the likelihood for the most general model with portfolio-specific breaks and parameters that is applied to each of the $i = 1, \ldots, N$ portfolios in turn is

$$p(r_i | X, \Theta_i, \tau_i) = \prod_{k_i=1}^{K_i+1} \prod_{t=\tau_{ik_i-1}+1}^{\tau_{ik_i}} p(r_{it} | X_{t-1}, \Theta_{it}),$$

$$= \prod_{k_i=1}^{K_i+1} (2\pi \sigma^2_{ik_i})^{-l_{ik_i}/2} \exp \left[ -\frac{1}{2} \sum_{k_i=1}^{K_i+1} \sum_{t=\tau_{ik_i-1}+1}^{\tau_{ik_i}} \frac{(r_{it} - X_{t-1}'\theta_{ik_i})^2}{\sigma^2_{ik_i}} \right],$$

(A.1)

in which $r_i$ denotes the $(T \times 1)$ vector of excess returns on the $i$th portfolio, and $X$ denotes the $(\kappa \times T)$ matrix which combines a unit vector with the observations on the predictor.\textsuperscript{21}

The likelihood function for the model with pooled breaks and unit-specific parameters is

$$p(r | X, \Theta, \tau) = \prod_{i=1}^{N} \prod_{k=1}^{K+1} \prod_{t=\tau_{k-1}+1}^{\tau_k} p(r_{it} | X_{t-1}, \Theta_{it}),$$

$$= \left( \prod_{i=1}^{N} \prod_{k=1}^{K+1} (2\pi \sigma^2_{ik})^{-l_{ik}/2} \exp \left[ -\frac{1}{2} \sum_{i=1}^{N} \sum_{k=1}^{K+1} \sum_{t=\tau_{k-1}+1}^{\tau_k} \frac{(r_{it} - X_{t-1}'\theta_{ik})^2}{\sigma^2_{ik}} \right] \right)$$

(A.2)

in which the breakpoints $\tau$ and hence regimes $k = 1, \ldots, K + 1$ are now common to all portfolios. Further let $r = (r_1, \ldots, r_N)$, $\mu = (\mu_1, \ldots, \mu_N)$, $\beta = (\beta_1, \ldots, \beta_N)$, $\sigma^2 = (\sigma^2_1, \ldots, \sigma^2_N)$, $\theta = (\theta_1, \ldots, \theta_N)$, and $\Theta = (\Theta_1, \ldots, \Theta_N)$.

We now provide details of the prefiltering approach to account for cross-sectional dependencies.

\textsuperscript{21}Making the first of the $\kappa = 2$ covariates a unit vector results in the first element of $\theta_i = (\mu_i, \beta_i)$ being an estimate of the intercept $\mu_i$ for the $i$th portfolio.
Appendix B. Accounting for Cross-sectional Dependencies

To deal with the cross-sectional dependencies we implement the popular approach that assumes the errors have a latent factor structure allowing for any number of unobserved common factors that induce cross-sectionally correlated errors. Following Pesaran (2006) we assume a setup:

\[ \tilde{r}_{it} = \beta'_{ik} \tilde{X}_{t-1} + \tilde{\epsilon}_t, \quad i = 1, \ldots, N, \quad t = \tau_{k-1} + 1, \ldots, \tau_k, \quad k = 1, \ldots, K + 1, \]

\[ \tilde{\epsilon}_t = \phi_{i1} f_{t-1} + \phi_{i2} f_t + \nu_t \]  

(B.1)
in which the errors \( \tilde{\epsilon}_t \) consist of not only the idiosyncratic errors \( \nu_t \) but also two common factors that we do not observe \( f_{t-1} \) and \( f_t \) and their corresponding factor loadings \( \phi_{i1} \) and \( \phi_{i2} \). There is no correlation between \( \nu_t \) and \( \tilde{X}_{t-1} \), but cross-sectional correlations may arise from the factors \( \tilde{X}_{t-1} = \Phi f_{t-1} + v_{t-1} \) 

(B.2)
in which the factor loading is denoted \( \Phi \) and the disturbance is denoted \( v_{t-1} \). Not accounting for such correlations can compromise inference; for example OLS estimation of (B.1) may be inconsistent. It is asymptotically valid to filter the data using cross-sectional averages of the dependent variable over time that proxy for any unobserved common factors (Pesaran 2006). This approach is particularly suited to our framework since Baltagi et al. (2016) shows that the approach remains asymptotically valid in the presence of multiple structural breaks as long as they are common, that is, any break hits every portfolio at the same date which we assume throughout. In Section 6.1 we explore whether this common break assumption is reasonable.

Letting \( C_{ik} = (\phi_{i1} + \Phi \beta'_{ik} \phi_{i2}) \) and \( u_{it} = \nu_{it} + \beta'_{ik} v_{t-1} \), we have

\[ \tilde{r}_{it} = C_{ik} \left( f_{t-1} f_t \right)' + u_{it}, \quad k = 1, \ldots, K + 1, \quad t = \tau_{k-1} + 1, \ldots, \tau_k \]  

(B.3)
in which \( C_{ik} \) also permanently shifts following a detected break. Further, let \( \bar{r}_t = \sum_{i=1}^{N} w_i \tilde{r}_{it} \) and \( \bar{u}_t = \sum_{i=1}^{N} w_i \tilde{u}_{it} \) such that:

\[ \bar{r}_t = \bar{C}_k \left( f_{t-1} f_t \right)' + \bar{u}_t, \quad k = 1, \ldots, K + 1, \quad t = \tau_{k-1} + 1, \ldots, \tau_k. \]  

(B.4)

\[ ^{22}\text{For expositional ease we have assumed the first element of } \tilde{X}_t \text{ is a unit vector such that the intercept term denoted } \mu_{it} \text{ in equation (2) is simply the first element of } \beta'_{it} \text{ here.} \]

\[ ^{23}\text{We use data on the number of firms in each portfolio and average firm size at each time period from Ken French’s website to construct the value weights } w_t \text{ for } t = 1, \ldots, T. \]
For the $i$th portfolio we can present (B.1) as

$$\tilde{R}_i = \beta_i' \tilde{X} + \tilde{\epsilon}_i$$  \hspace{1cm} (B.5)

in which $\tilde{R}_i = (\tilde{r}_{i2}, \ldots, \tilde{r}_{iT})$, $X = (X_1, \ldots, X_{T-1})$ and $\tilde{\epsilon}_i = (\tilde{\epsilon}_{i2}, \ldots, \tilde{\epsilon}_{iT})$. Further let $\tilde{r} = (\tilde{r}_2, \ldots, \tilde{r}_T)$ denote the excess returns on the market portfolio and define the orthogonal projection matrix $M_r = I_{T-1} - \tilde{r}(\tilde{r}'\tilde{r})^{-1}\tilde{r}'$ such that premultiplying (B.5) by $M_r$ obtains

$$R_i = \mu_i + \beta_i' X + \epsilon_i, \quad i = 1, \ldots, N$$  \hspace{1cm} (B.6)

in which $\epsilon_i = M_r F^{-1} \phi_{i1} + M_r F \phi_{i2} + M_r \nu_{i1}$, $R_i = M_r \tilde{R}_i$, $X = M_r \tilde{X}$, $F = (f_2, \ldots, f_T)$, and $F_{-1} = (f_1, \ldots, f_{T-1})$. As $N \to \infty$ both $M_r F_{-1} \phi_{i1}$ and $M_r F \phi_{i2} \to 0$ and thus performing our estimation procedure (described in Appendix C) on the transformed data is asymptotically valid since we effectively treat $\epsilon_i$ as if it were $M_r \nu_{i1}$.\footnote{We have explicitly introduced the intercept as a separate term here for consistency with the remainder of the paper, thus the first element of $X$ is no longer a unit vector.}

We now provide details of the prior distributions used by our return prediction models.

**Appendix C. Priors**

**Appendix C.0.1. Prior on the Regime Durations**

*Unit-specific Breaks:* Chib (1998)'s method restricts the regime duration of each time-series in the cross-section to follow a geometric prior distribution

$$p(l_{ik_i} \mid p_{k_i}, K) = p_{k_i}^{l_{ik_i} - 1}(1 - p_{k_i}), \quad k_i = 1, \ldots, K_i + 1$$ \hspace{1cm} (C.1)

in which the prior nontransition probability $p_{k_i}$ follows a conjugate beta distribution

$$p(p_{k_i}) = \frac{\Gamma(g + h)}{\Gamma(g)\Gamma(h)} p_{k_i}^{g-1}(1 - p_{k_i})^{h-1}, \quad k_i = 1, \ldots, K_i$$ \hspace{1cm} (C.2)

and $p_{k_i}$ by construction. We therefore have

$$p(\tau) = \prod_{k_i=1}^{K_i+1} p(l_{ik_i} \mid p_{k_i}, K) \left( \prod_{k_i=1}^{K_i} p(p_{k_i}) \right).$$ \hspace{1cm} (C.3)
Pooled Breaks: Koop and Potter (2007) note that such a monotonically decreasing geometric prior on the regime durations enforced by Chib (1998)’s algorithm may be unrealistic and therefore suggest specifying a Poisson distribution instead. In the panel break model we develop the regime durations have a Poisson prior distribution

\[ p(l_k \mid \lambda_k, K) = \frac{\lambda_k^{l_k} e^{-\lambda_k}}{l_k!}, \quad k = 1, \ldots, K + 1, \]  

(C.4)

where the hyperparameter \( \lambda_k \), which corresponds to the expected prior duration of regime \( k \), has a conjugate Gamma prior distribution

\[ p(\lambda_k) = \frac{d^c}{\Gamma(c)} \lambda_k^{c-1} e^{-d\lambda_k}, \quad k = 1, \ldots, K + 1, \]  

(C.5)

in which \( c \) and \( d \) are the hyperparameters of \( \lambda = (\lambda_1, \ldots, \lambda_{K+1}) \). Having marginalised and discarded \( \lambda \) the prior on the breakpoints is therefore

\[ p(\tau) = \prod_{k=1}^{K+1} \frac{1}{l_k! (d + 1)^{c+l_k} \Gamma(c)} \]  

(C.6)

Appendix C.0.2. Priors on Parameters \( \beta \) and \( \sigma \)

The estimation of the panel break model is simplified by specifying conjugate priors on the error-term variances \( \sigma^2 \) and on the regression coefficients \( \theta \) conditional on the error-term variances \( \sigma^2 \). For consistency we use the same prior specification for the unit-specific break and parameter model.

Unit-specific Breaks: For the \( i \)th portfolio we specify a conjugate inverse gamma prior over the error-term variances

\[ p(\sigma_{ik_i}^2) = \frac{b^a}{\Gamma(a)} \sigma_{ik_i}^{2-(a+1)} e^{-b \sigma_{ik_i}^2}, \quad k_i = 1, \ldots, K_i + 1, \]  

(C.7)

and a conjugate normal prior with zero mean over the intercept and slope coefficient \( \theta_i = (\mu_i, \beta_i) \) conditional on the error-term variances \( \sigma_i^2 \)

\[ p(\theta_{ik_i} \mid \sigma_{ik_i}^2) = 2\pi^{-\kappa/2} (\sigma_{ik_i}^2)^{-\kappa/2} \left| V_{\beta} \right|^{-1/2} e^{\frac{1}{2\sigma_{ik_i}^2} (\theta_{ik_i}^T V_{\beta}^{-1} \theta_{ik_i})}, \]  

\[ V_{\beta} = \begin{pmatrix} \sigma_{\mu}^2 & 0 \\ 0 & \sigma_{\eta}^2 / \sigma_{z}^2 \end{pmatrix}. \]  

(C.8)
We therefore have

\[ p(\sigma_i^2) = \prod_{k_i=1}^{K_i+1} p(\sigma_{ik_i}^2), \]

\[ p(\theta_i | \sigma_i^2) = \prod_{k_i=1}^{K_i+1} p(\theta_{ik_i} | \sigma_{ik_i}^2). \]  \hspace{1cm} \text{(C.9)}

**Pooled Breaks and Unit-specific Parameters:** For the model with pooled breaks and unit-specific parameters we specify an inverse gamma prior over the regime-specific variances for \( i = 1, \ldots, N \)

\[ p(\sigma_{ik}^2) = \frac{b^a}{\Gamma(a)} \sigma_{ik}^{2-a} \exp\left( -\frac{b}{\sigma_{ik}^2} \right), \quad k = 1, \ldots, K+1. \]  \hspace{1cm} \text{(C.10)}

and a normal prior with zero mean is placed over the regression coefficients conditional on the variances

\[ p(\theta_{ik} | \sigma_{ik}^2) = 2\pi^{-\kappa/2}(\sigma_{ik}^2)^{-\kappa/2} \left| V_\beta \right|^{-1/2} \exp\left( -\frac{1}{2\sigma_{ik}^2} \theta_{ik}'V_\beta^{-1}\theta_{ik} \right), k = 1, \ldots, K+1 \]  \hspace{1cm} \text{(C.11)}

We therefore have

\[ p(\sigma^2) = \prod_{i=1}^{N} \prod_{k=1}^{K+1} p(\sigma_{ik}^2), \]

\[ p(\theta | \sigma^2) = \prod_{i=1}^{N} \prod_{k=1}^{K+1} p(\theta_{ik} | \sigma_{ik}^2). \]  \hspace{1cm} \text{(C.12)}

**Pooled Breaks and Parameters** For the model with pooled breaks and parameters we specify an inverse gamma prior over the regime-specific variances

\[ p(\sigma_k^2) = \frac{b^a}{\Gamma(a)} \sigma_k^{2-a} \exp\left( -\frac{b}{\sigma_k^2} \right), \quad k = 1, \ldots, K+1, \]  \hspace{1cm} \text{(C.13)}

and a normal prior with zero mean is placed over the regression coefficients conditional on the variances

\[ p(\theta_k | \sigma_k^2) = 2\pi^{-\kappa/2}(\sigma_k^2)^{-\kappa/2} \left| V_\beta \right|^{-1/2} \exp\left( -\frac{1}{2\sigma_k^2} \theta_k'V_\beta^{-1}\theta_k \right), k = 1, \ldots, K+1. \]  \hspace{1cm} \text{(C.14)}
We therefore have\footnote{The prior specifications for the two partial break models are the same as in the pooled break and parameter model except the prior is being placed over the constant error-term variance or intercept, respectively, rather than over the regime-specific values. For simplicity we do not explicitly detail these priors.}

\[
p(\sigma^2) = \prod_{k=1}^{K+1} p(\sigma_k^2),
\]

\[
p(\theta | \sigma^2) = \prod_{k=1}^{K+1} p(\theta_k | \sigma_k^2).
\]

\hspace{1cm} \text{(C.15)}

\textit{Appendix C.1. Prior Elicitation}

For the empirical application we assume that a break occurs approximately every decade for the heterogeneous panel break model we develop and the benchmark time-series break model. This is in line with findings in earlier studies such as Pástor and Stambaugh (2001). Specifically, we set our hyperparameter values as \( d = h = 2 \) and \( c = g = 240 \) to give a prior expected regime duration of 120 periods. We further set \( a = 2 \) and \( b = 0.0049 \) to give a prior expected error-term variance equal to 0.0049 which is the average of the variance of the excess returns across all the portfolios using the full sample. The choice of hyperparameter values for \( \sigma^2_\mu \) and \( \sigma^2_\eta \), and the scaling of \( \sigma^2_\eta \) with the empirical variance of the predictive variable using the full sample available at the time each forecast is made \( \sigma^2_x \), have been explained in Section 2.5.

\textit{Appendix C.2. Posterior Distribution}

Inference is performed on the posterior distribution which combines information in the data with prior information supplied by the user. We now explicitly detail the posterior distributions of the different models we consider.
**Unit-specific Breaks and Parameters:** The posterior distribution of the model with unit-specific breaks and parameters is

\[
p(\Theta_i | r_i, X, \tau_i) = \left( \prod_{k_i=1}^{K_i+1} (\sigma_{ik_i}^2)^{-((l_{ik_i}+\kappa)/2+\alpha+1)}(2\pi)^{-\alpha(l_{ik_i})/2} \right) \times \exp \left[ -\frac{1}{2} \sum_{k_i=1}^{K_i+1} \left( \frac{\theta_{ik_i}^\prime V_\beta^{-1} \theta_{ik_i} + 2b + (r_{ik_i}^\prime r_{ik_i} - r_{ik_i}^\prime X_{k_i}^\prime \theta_{ik_i} - \theta_{ik_i}^\prime X_{k_i} r_{ik_i} + \theta_{ik_i}^\prime X_{k_i} X_{k_i}' \theta_{ik_i})}{\sigma_{ik_i}^2} \right) \right] ,
\]

(C.16)

in which we denote the \((l_{ik_i} \times 1)\) vector \(r_{ik_i} = (r_{i\tau_{k_i}-1} \ldots, r_{i\tau_{k_i}})\) and the \((\kappa \times l_{ik_i})\) matrix

\[
X_{k_i}^{(\kappa \times l_{ik_i})} = \begin{pmatrix} X_{1\tau_{k_i}-1+1} & \ldots & X_{1\tau_{k_i}} \\ \vdots & \ddots & \vdots \\ X_{\kappa\tau_{k_i}-1+1} & \ldots & X_{\kappa\tau_{k_i}} \end{pmatrix}.
\]

**Pooled Breaks and Unit-specific Parameters** The posterior distribution of the model with pooled breaks and unit-specific parameters is

\[
p(\Theta | r, X, \tau) = \left( \prod_{i=1}^{N} \prod_{k=1}^{K+1} (\sigma_{ik}^2)^{-((l_k+\kappa)/2+\alpha+1)}(2\pi)^{-\alpha(l_k)/2} \right) \times \exp \left[ -\frac{1}{2} \sum_{i=1}^{N} \sum_{k=1}^{K+1} \left( \frac{\theta_{ik}^\prime V_\beta^{-1} \theta_{ik} + 2b + (r_{ik}^\prime r_{ik} - r_{ik}^\prime X_{k}^\prime \theta_{ik} - \theta_{ik}^\prime X_{k} r_{ik} + \theta_{ik}^\prime X_{k} X_{k}' \theta_{ik})}{\sigma_{ik}^2} \right) \right] ,
\]

(C.17)

in which we denote the \((l_k \times 1)\) vector \(r_{ik} = (r_{i\tau_{k-1}+1} \ldots, r_{i\tau_{k}})\), and the \((\kappa \times l_k)\) matrix

\[
X_k^{(\kappa \times l_k)} = \begin{pmatrix} X_{1\tau_{k-1}+1} & \ldots & X_{1\tau_k} \\ \vdots & \ddots & \vdots \\ X_{\kappa\tau_{k-1}+1} & \ldots & X_{\kappa\tau_k} \end{pmatrix}.
\]

In order to marginalise \(\theta\) from the posterior let us define

\[
\Sigma_k^{-1} = V_\beta^{-1} + X_k X_k', \quad k = 1, \ldots, K + 1 ,
\]

(C.18)

and

\[
\rho_{ik} = \Sigma_k X_k r_{ik}, \quad k = 1, \ldots, K + 1, \quad i = 1, \ldots, N
\]

(C.19)
and let the \((N \times 1)\) vectors \(\rho_1 = (\rho_{1,1}, \ldots, \rho_{N,1})', \ldots, \rho_{K+1} = (\rho_{1,K+1}, \ldots, \rho_{N,K+1})',\) and let \(\bm{\mu} = (\mu_1, \ldots, \mu_{K+1})\). Let \(\Sigma^{-1} = (\Sigma^{-1}_1, \ldots, \Sigma^{-1}_{K+1})\). We now obtain

\[
p(\Theta \mid \bm{r}, \bm{X}, \tau) = \left( \prod_{i=1}^{N} \prod_{k=1}^{K+1} \left( \sigma^2_{ik} \right)^{-(l_k + a_1 + 1)/2} (2\pi)^{-l_k/2} V_{\beta}^{-1/2} \frac{\Gamma(a)}{\Gamma(a)} \right) \times \exp \left[ -\frac{1}{2} \sum_{i=1}^{N} \sum_{k=1}^{K+1} \left( \frac{\theta_{ik}' \Sigma_k^{-1} \theta_{ik}}{\sigma^2_{ik}} \right) \right],
\]

in which \(\theta_{ik}' \Sigma_k^{-1} \theta_{ik}\) in (C.20) corresponds to \(\theta_{ik}' V^{-1}_{\beta} \theta_{ik}\) and \(\theta_{ik}' X_k X_k' \theta_{ik}\) in (C.17), \(\theta_{ik}' \Sigma_k^{-1} \rho_{ik}\) corresponds to \(\theta_{ik}' X_k r_{ik}\), and \(\rho_{ik}' \Sigma_k^{-1} \theta_{ik}\) corresponds to \(r_{ik}' X_k' \theta_{ik}\).

Multiplying and dividing by \(2\pi^{k/2} |\Sigma_k|^{1/2} (\sigma^2_{ik})^{k/2}\) enables \(\theta_{ik}\) to be marginalised from the posterior since it has a multivariate normal (full conditional) distribution with mean \(\rho_{ik}\) and covariance matrix \(\Sigma_k \sigma^2_{ik}\) for \(k = 1, \ldots, K + 1\) and \(i = 1, \ldots, N\) that will integrate to 1 leaving

\[
p(\sigma^2 \mid \bm{r}, \bm{X}, \tau) = \left( \prod_{i=1}^{N} \prod_{k=1}^{K+1} \left( \sigma^2_{ik} \right)^{-l_k/2} (2\pi)^{-l_k/2} V_{\beta}^{-1/2} \frac{\Gamma(a)}{\Gamma(a)} |\Sigma_k|^{1/2} \right) \times \exp \left[ -\frac{1}{2} \sum_{i=1}^{N} \sum_{k=1}^{K+1} \left( \frac{\rho_{ik}' \Sigma_k^{-1} \rho_{ik}}{\sigma^2_{ik}} \right) \right],
\]

Let us further define

\[
\tilde{a}_k = a + (l_k)/2, \quad k = 1, \ldots, K + 1,
\]

and

\[
\bar{b}_{ik} = \frac{1}{2} \left( 2b + r_{ik}' r_{ik} - \rho_{ik}' \Sigma_k^{-1} \rho_{ik} \right), \quad k = 1, \ldots, K + 1, \quad i = 1, \ldots, N,
\]

and let \(\tilde{a} = (\tilde{a}_1, \ldots, \tilde{a}_{K+1})\). Let the \((N \times 1)\) vectors \(\tilde{b}_1 = (\tilde{b}_{1,1}, \ldots, \tilde{b}_{1,N})', \ldots, \tilde{b}_{K+1} = (\tilde{b}_{K+1,1}, \ldots, \tilde{b}_{K+1,N})',\) and \(\tilde{b} = (\tilde{b}_1, \ldots, \tilde{b}_{K+1})\). Multiplying and dividing by \(\Gamma(\tilde{a}_k)/\tilde{b}_{ik}\) enables us to marginalise \(\sigma^2\) from the posterior since \(\sigma^2_{ik}\) has an inverse gamma (full conditional) distribution with hyperparameters \(\tilde{a}_k\) and \(\bar{b}_{ik}\) for \(k = 1, \ldots, K + 1\) and
\[ p(r | X, \tau) = \prod_{i=1}^{N} \prod_{k=1}^{K+1} (2\pi)^{-l_k/2} \left| \Sigma_k \right|^{1/2} \left| V_\beta \right|^{1/2} \Gamma(\alpha) \Gamma(\tilde{\alpha}_k) \delta_{ik}^{\alpha_b} \delta_{ik}^{\tilde{\beta}} \]

(C.24)

Appendix D. Estimation of the Model

This appendix provides details of the procedures used to estimate the different models considered in the paper. The model with unit-specific breaks and parameters is repeatedly estimated for each time-series in the cross-section using the multiple breakpoint model of Chib (1998). This procedure estimates a series of models each with a different number of breaks and subsequently uses the marginal likelihood approach of Chib (1995) to derive the posterior model probabilities and determine the optimal number of breaks. Given the popularity of Chib (1998)'s algorithm in economics and finance along with the desire to save space we do not present the details of the algorithm here.

In contrast the models with common breaks analyse the entire cross-section at once using an alternative estimation procedure that introduces the number of breaks as a parameter in the model and performs inference over this parameter by jumping between different numbers of breaks and thereby the proportion of the Markov chain Monte Carlo run that is spent at each number of breaks approximates the posterior model probabilities (Green 1995). Our estimation approach has a range of desirable properties relative to Chib (1998); we refer the reader to Smith and Timmermann (2017) for a thorough discussion. Specifically we simulate the breakpoint vector \( \tau \) in two steps. First, a global movement is provided by attempting to either add or remove a breakpoint on each sweep of the MCMC run. Second, to ensure the estimated breakpoint locations converge to their true values all that is required is a small perturbation of each breakpoint delivered by a random-walk Metropolis-Hastings step. Finally the parameters can be sampled from the full conditionals. We now formally explain estimation of the breakpoint vector and the parameter vector for all the models we consider.

Appendix D.1. Estimating the Parameter Vector

Given the estimated number of breakpoints and their locations it is straightforward to estimate the parameters from their full conditional distributions using the Gibbs
sampler. We now detail the full conditional distributions from which the parameters of each of the models we consider are sampled.

For the model with pooled breaks and unit-specific parameters the full conditional distributions are

\[
\begin{align*}
\sigma_{ik}^2 | \cdot & \sim IG(\tilde{a}_k, \tilde{b}_{ik}), \quad k = 1, \ldots, K + 1, \quad i = 1, \ldots, N, \\
\theta_{ik} | \cdot & \sim MVN(\rho_{ik}, \Sigma_k \sigma_{ik}^2), \quad k = 1, \ldots, K + 1, \quad i = 1, \ldots, N
\end{align*}
\]

(D.1)
in which we have already computed the corresponding values of \(\mu, \Sigma, \tilde{a},\) and \(\tilde{b}\) during estimation of the breakpoints.

Appendix D.2. Estimating the breakpoint Locations

This section explains how we estimate the location of the breakpoints in the cases with unit-specific and common break dates.

Since varying the number of breakpoints provides the sampler with a global movement all that is required to keep the breakpoint locations true is a local adjustment in the form of a random-walk Metropolis-Hastings step. For \(k \in (1, K)\) each breakpoint \(\tau_k\) is perturbed by a discrete number \(u\) that is sampled uniformly from the interval \([-s, s]\) to give the proposed breakpoint \(\tau_k^*\).\(^{27}\) Let \(\tau^*\) denote the proposed breakpoint vector which is equal to the current breakpoint vector \(\tau\) except \(\tau_k\) is replaced by \(\tau_k^*\). If \(\tau_k^* \in \tau\) the proposal is immediately rejected otherwise it is accepted with probability \(\min(1, \alpha)\) having computed \(l_{k^*} = \tau_{k^*} - \tau_{k-1}\) and \(l_{k^*+1} = \tau_{k+1} - \tau_{k^*}\), \(\tilde{a}_{k^*}\) and \(\tilde{a}_{k^*+1}\) using (C.22), \(\Sigma_{k^*+1}^{-1}\) using (C.18) and, for \(i = 1, \ldots, N, \Sigma_k^{-1}, \rho_{ik^*}\) and \(\rho_{ik^*+1}\) using (C.19), and \(\tilde{b}_{ik}\) and \(\tilde{b}_{ik+1}\) using (C.23).

Using (C.6) and (C.24) we compute \(\alpha\) as

\[
\alpha = \frac{\Gamma(l_{k^*} + c) \Gamma(l_{k^*+1} + c)}{\Gamma(l_k + c) \Gamma(l_{k+1} + c)} \frac{l_k! l_{k+1}!}{l_{k^*}! l_{k^*+1}!} \prod_{i=1}^{N} \frac{|\Sigma_k|^{1/2} |\Sigma_{k^*+1}|^{1/2} \Gamma(\tilde{a}_{k^*}) \Gamma(\tilde{a}_{k^*+1}) \tilde{b}_{ik} \tilde{b}_{ik+1}}{|\Sigma_k|^{1/2} |\Sigma_{k+1}|^{1/2} \Gamma(\tilde{a}_{k}) \Gamma(\tilde{a}_{k+1}) \tilde{b}_{ik} \tilde{b}_{ik+1}}
\]

(D.2)
in which canceling is observed due to constant terms and because of the equality \(l_{k^*} + l_{k^*+1} = l_k + l_{k+1}\), while all but the two regimes separated by the breakpoint that is being perturbed remain unchanged and thus do not enter the acceptance calculation.

\(^{27}\)We set \(s\) equal to 1 to achieve the desired acceptance ratio of approximately 25%.
If \( \tau_{k^*} \) is rejected we simply discard it and then attempt to perturb the next breakpoint. If accepted we replace \( \tau_k \) with \( \tau_{k^*} \) and replace the existing values of \( l, \tilde{a}, \tilde{b}, \mu, \) and \( \Sigma^{-1} \) for the current regimes \( k \) and \( k+1 \) with the corresponding accepted proposed values for \( i = 1, \ldots, N \) in regimes \( k^* \) and \( k^*+1 \).

Appendix D.3. Estimating the Number of breakpoints

We adopt the reversible jump Markov chain Monte Carlo approach that includes the number of breaks \( K \) as a parameter in the model and explores both the model and parameter space jointly by ‘jumping’ between different numbers of breaks and the proportion of time spent at each number of breaks is equal to the posterior model probabilities. The choice of conjugate priors on the regression parameters \( \mu, \beta, \) and \( \sigma \) allowed us to marginalise them from the posterior and thereby explore the model space alone reducing the complexity of the algorithm.

To estimate the number of breakpoints we begin the sampler at a sensible number of breakpoints. On each sweep of the Markov chain Monte Carlo run we attempt with equal probability to either add (birth move) or remove a breakpoint (death move). We then compute the acceptance probability that ensures detailed balance is maintained across the entire parameter space including the number of breaks. If the move is accepted the breakpoint vector is updated otherwise the proposed breakpoint vector is discarded.

**Birth Move:** With probability \( b_K = 0.5 \) a birth move is entered.\(^{28}\) This move attempts to increase \( K \) to \( K + 1 \) and hence introduce a new breakpoint denoted \( \tau_{k^*} \) that is sampled uniformly from the entire time series sample \( \tau_{k^*} \sim U[1, T] \). If an existing breakpoint is proposed, that is, if \( \tau_{k^*} \in \tau \), the move is immediately rejected. Otherwise let \( \tau^* \) denote the proposed breakpoint vector which consists of \( \tau \) and \( \tau_{k^*} \). Let \( k^c \) denote the existing regime we are attempting to split, i.e. \( \tau_{k^c-1} < \tau_{k^*} < \tau_{k^c} \), and let \( k^* \) and \( k^*+1 \) denote the two new regimes we are proposing to introduce. Regime \( k^* \) contains the observations \( \tau_{k^c-1} + 1, \ldots, \tau_{k^*} \) and hence \( l_{k^*} = \tau_{k^*} - \tau_{k^c-1} \). Regime \( k^*+1 \) contains the observations \( \tau_{k^c} + 1, \ldots, \tau_{k^c} \) and hence \( l_{k^*+1} = \tau_{k^c} - \tau_{k^*} \). We now compute \( \tilde{a}_{k^*} \) and \( \tilde{a}_{k^*+1} \) using (C.22), \( \Sigma_{k^*}^{-1} \) and \( \Sigma_{k^*+1}^{-1} \) using (C.18) and, for \( i = 1, \ldots, N, \rho_{ik^*} \) and \( \rho_{ik^*+1} \) using (C.19), and \( \tilde{b}_{k^*} \) and \( \tilde{b}_{k^*+1} \) using (C.23). Since the parameter vector \( \Theta = (\mu, \beta, \sigma^2) \) has been marginalised from the posterior there is no

\(^{28}\)Unless the sampler is at \( K = T - 1 \), which corresponds to a break occurring every period, in which \( b_{T-1} = 0 \).
need to propose their values here.

The birth move is accepted with probability equal to $\min(1, \alpha)$ whereby $\alpha$ is equal to

$$\alpha = \frac{p(r | X, \tau^*)}{p(r | X, \tau)} \times \frac{p(\tau^*)}{p(\tau)} \times \frac{T}{K + 1} \times \frac{2}{2},$$

(D.3)

in which the final term $2/2$ cancels due to the use of an equal probability of a birth and death move being selected ($b_k = 1 - b_k = 1/2$), $T$ corresponds to the uniform sampling of $k^*$ on the interval $[1, T]$, and $(K + 1)^{-1}$ corresponds to the uniform sampling of an existing breakpoint to return to $K$ breakpoints if we had $K + 1$.\(^{29}\) The remaining terms are calculated using (C.6) and (C.24). Specifically $\alpha$ is equal to

$$\alpha = \frac{d^c l_k c! 1 \Gamma(l_{k^*+1} + c) \Gamma(l_{k^*} + c)}{\Gamma(c) l_{k^*}! (l_{k^*+1})! \Gamma(l_{k^*} + c)} \frac{T}{K + 1} \frac{b a N}{\Gamma(a) N} \times \prod_{i=1}^{N} \frac{\tilde{a}_{k^*+1}}{\tilde{a}_{k^*}} \frac{\Gamma(\tilde{a}_{k^*}) \Gamma(\tilde{a}_{k^*+1})}{\Gamma(\tilde{a}_{k^*})} \frac{|\Sigma_{k^*+1}|^{1/2}}{|\Sigma_{k^*}|^{1/2}} \frac{|\Sigma_{k^*}|^{1/2}}{|\Sigma_{k^*}|^{1/2}},$$

(D.4)

in which all but the existing regime we are proposing to split remain unchanged and thus do not enter into the acceptance probability.

If the move is rejected the proposals are discarded otherwise $\tau$ is replaced by $\tau^*$, while the corresponding values of $l, \Sigma^{-1}, \rho, \tilde{a},$ and $\tilde{b}$ are updated by removing their values for the existing regime $k^c$ and adding the accepted proposed values for the two new regimes $k^*$ and $k^* + 1$.

**Death Move:** With probability $d_k = 1 - b_k = 0.5$ a death move is entered.\(^{30}\) This move attempts to remove an existing breakpoint and hence reduce $K$ to $K - 1$. The existing breakpoint $\tau_{k^c}$ is sampled uniformly from the existing breakpoint vector $\tau_{k^c} \sim U[\tau_1, \tau_K]$. Let $\tau^*$ denote the proposed breakpoint vector which is equal to $\tau$ with $\tau_{k^c}$ removed. We are proposing to replace two existing regimes denoted $k^c$ and $k^c + 1$ with one new regime denoted $k^*$. The length of the proposed regime is simply the sum of the two existing regimes $l_{k^*} = l_{k^c} + l_{k^c+1}$. In similar fashion to the birth move, we now compute $\tilde{a}_{k^*}$ using (C.22), $\Sigma_{k^*}^{-1}$ using (C.18) and, for $i = 1, \ldots, N$, $\rho_{ik^*}$ using (C.19), and $\tilde{b}_{ik^*}$ using (C.23).

We accept the death move with probability equal to $\min(1, \alpha)$ whereby $\alpha$ is equal to

---

\(^{29}\)This final term enters the acceptance ratio to ensure every move is ‘reversible’ otherwise the algorithm might move to a part of the parameter space from which it cannot return and could thus stop mixing.

\(^{30}\)Unless the sampler is at $K = 0$ in which $d_0 = 0$ since there is no existing breakpoint to remove.
to
\[ \alpha = \frac{p(r \mid X, \tau^*)}{p(r \mid X, \tau)} \times \frac{p(\tau^*)}{p(\tau)} \times \frac{K}{T} \times \frac{2}{2}, \tag{D.5} \]
in which the final term cancels due to the use of an equal probability for both move types, \( K \) corresponds to the uniform sampling of \( k^c \) from the existing set of \( K \) breakpoints, and \( T^{-1} \) corresponds to the uniform sampling of a new breakpoint from the interval \( U[1, T] \) if we were at \( K - 1 \) breakpoints and attempting a birth move to \( K \) breakpoints. The remaining terms are calculated using (C.6) and (C.24). Specifically \( \alpha \) is equal to
\[ \alpha = \frac{l_{k^c} l_{k^e+1}}{l_{k^e}} \frac{\Gamma(l_{k^e} + c)}{\Gamma(l_{k^e} + c) \Gamma(l_{k^e+1} + c)} \frac{(d + 1)^c}{d^c} \Gamma(c) K \]
\[ \times \prod_{i=1}^{N} \frac{\Gamma(\tilde{a}_{k^c})}{\Gamma(\tilde{a}_{k^e})} \frac{\tilde{b}^{a_{k^e}}_{k^e}}{\tilde{b}^{a_{k^e+1}}_{k^e+1}} \frac{|\Sigma_{k^e}|^{1/2} V_\beta^{1/2}}{|\Sigma_{k^e+1}|^{1/2} V_\beta^{1/2}} \frac{\tilde{a}_{k^e+1}}{\tilde{a}_{k^e+1}} \frac{\Gamma(a)}{\Gamma(a)} \tag{D.6} \]
If the move is rejected the proposals are discarded otherwise \( \tau \) is replaced by \( \tau^* \), while \( l, \Sigma^{-1}, \mu, \tilde{a}, \) and \( \tilde{b} \) are updated by removing their values for the existing regimes \( k^e \) and \( k^e + 1 \) and adding the accepted proposed values for the new regime \( k^* \).
Table 1: Statistical significance of improved forecasting power

<table>
<thead>
<tr>
<th>Predictor</th>
<th>DM</th>
<th>CW</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$t &lt; -1.64$</td>
<td>$-1.64 &lt; t &lt; 0$</td>
</tr>
<tr>
<td><strong>No break panel</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dp</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>tbl</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>tms</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>dfs</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td><strong>Industry prevailing mean</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dp</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>tbl</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>tms</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>dfs</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td><strong>Time series breakpoint</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dp</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>tbl</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>tms</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>dfs</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td><strong>Pooled panel</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dp</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>tbl</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>tms</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>dfs</td>
<td>8</td>
<td>5</td>
</tr>
</tbody>
</table>

Table 1: Statistical significance of forecasts. This table reports the statistical significance of the improved forecasting power of the method we develop relative to the heterogeneous panel model with no breaks (No breakpoint), the industry prevailing mean, the time series model with breaks applied to each portfolio in turn (Time series breakpoint), and the panel break model with both pooled parameters and breaks (Pooled panel) when forecasting with the dividend-price ratio (dp), the treasury-bill rate (tbl), the term spread (tms), and the default spread (dfs). Significance is evaluated using the Diebold-Mariano test (DM) and the procedure of Clark and West (2007) (CW). For each procedure the table displays the number of the portfolios for which our method produces significantly worse, insignificantly worse, insignificantly better, and significantly better forecasts at the 10% level. † indicates the particular bin in which the $t$-statistic for the market portfolio lies.
Table 2: Magnitude of break by portfolio

<table>
<thead>
<tr>
<th>Industry</th>
<th>Size of break rank</th>
<th>MSFD</th>
<th>Utility Gain rank</th>
<th>Utility gain</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Upper quartile</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>telem</td>
<td>1</td>
<td>0.0220</td>
<td>1</td>
<td>2.48</td>
</tr>
<tr>
<td>util</td>
<td>2</td>
<td>0.0172</td>
<td>13</td>
<td>1.63</td>
</tr>
<tr>
<td>oil</td>
<td>3</td>
<td>0.0148</td>
<td>7</td>
<td>1.79</td>
</tr>
<tr>
<td>buseq</td>
<td>4</td>
<td>0.0144</td>
<td>6</td>
<td>1.94</td>
</tr>
<tr>
<td>fin</td>
<td>5</td>
<td>0.0142</td>
<td>2</td>
<td>2.40</td>
</tr>
<tr>
<td>hlth</td>
<td>6</td>
<td>0.0140</td>
<td>3</td>
<td>2.31</td>
</tr>
<tr>
<td>beer</td>
<td>7</td>
<td>0.0129</td>
<td>5</td>
<td>1.97</td>
</tr>
<tr>
<td>fabpr</td>
<td>24</td>
<td>0.0059</td>
<td>16</td>
<td>1.45</td>
</tr>
<tr>
<td>whlsl</td>
<td>25</td>
<td>0.0055</td>
<td>29</td>
<td>-0.34</td>
</tr>
<tr>
<td>textiles</td>
<td>26</td>
<td>0.0050</td>
<td>22</td>
<td>1.04</td>
</tr>
<tr>
<td>mines</td>
<td>27</td>
<td>0.0041</td>
<td>26</td>
<td>0.67</td>
</tr>
<tr>
<td>books</td>
<td>28</td>
<td>0.0033</td>
<td>28</td>
<td>0.15</td>
</tr>
<tr>
<td>meals</td>
<td>29</td>
<td>0.0029</td>
<td>30</td>
<td>-0.47</td>
</tr>
<tr>
<td>other</td>
<td>30</td>
<td>0.0024</td>
<td>27</td>
<td>0.61</td>
</tr>
</tbody>
</table>

Table 2: Magnitude of break by industry. This table lists in descending order the upper and lower quartile portfolios according to the magnitude of the total impact of breaks on their respective forecasts (with 1 denoting the largest impact). This magnitude is captured by the mean squared forecast difference (‘MSFD’) between the panel models with and without breaks. The table also reports the ranking of the utility gain, expressed as an annualised percentage, for a mean-variance investor with a risk aversion coefficient of three when forecasting with the panel break model relative to the panel model without breaks using the dividend-price ratio as the predictive variable.
### Table 3: Industry allocation

<table>
<thead>
<tr>
<th>Industry</th>
<th>Hist avg</th>
<th>No brk</th>
<th>ts</th>
<th>Pbrk</th>
</tr>
</thead>
<tbody>
<tr>
<td>food</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>beer</td>
<td>0.23</td>
<td>0.00</td>
<td>0.11</td>
<td>0.19</td>
</tr>
<tr>
<td>smoke</td>
<td>0.09</td>
<td>0.24</td>
<td>0.07</td>
<td>0.13</td>
</tr>
<tr>
<td>books</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.02</td>
</tr>
<tr>
<td>hlth</td>
<td>0.06</td>
<td>0.02</td>
<td>0.01</td>
<td>0.00</td>
</tr>
<tr>
<td>chems</td>
<td>0.14</td>
<td>0.06</td>
<td>0.13</td>
<td>0.04</td>
</tr>
<tr>
<td>txtls</td>
<td>0.00</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>elceq</td>
<td>0.02</td>
<td>0.01</td>
<td>0.00</td>
<td>0.01</td>
</tr>
<tr>
<td>autos</td>
<td>0.07</td>
<td>0.02</td>
<td>0.04</td>
<td>0.00</td>
</tr>
<tr>
<td>oil</td>
<td>0.04</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>telem</td>
<td>0.03</td>
<td>0.01</td>
<td>0.23</td>
<td>0.05</td>
</tr>
<tr>
<td>servs</td>
<td>0.16</td>
<td>0.45</td>
<td>0.37</td>
<td>0.42</td>
</tr>
<tr>
<td>buseq</td>
<td>0.11</td>
<td>0.17</td>
<td>0.01</td>
<td>0.07</td>
</tr>
<tr>
<td>paper</td>
<td>0.02</td>
<td>0.00</td>
<td>0.01</td>
<td>0.00</td>
</tr>
<tr>
<td>fin</td>
<td>0.00</td>
<td>0.00</td>
<td>0.01</td>
<td>0.05</td>
</tr>
</tbody>
</table>

**Table 3: Allocations between industries.** This table reports average allocation to each of the thirty industry portfolios across the out-of-sample period using the four competing models when predicting with the dividend-price ratio. We omit industries which have an allocation of less than 0.01 for each model. ‘Hist avg’ denotes the industry prevailing mean, ‘No brk’ denotes the heterogeneous panel model without breaks, ‘ts’ denotes the time series break model, and ‘Pbrk’ denotes the heterogeneous panel model with breaks.

### Table 4: Portfolio utility

<table>
<thead>
<tr>
<th>Predictor</th>
<th>hist avg</th>
<th>no brk</th>
<th>ts</th>
<th>hist avg</th>
<th>no brk</th>
<th>ts</th>
</tr>
</thead>
<tbody>
<tr>
<td>dp</td>
<td>2.19</td>
<td>2.02</td>
<td>1.97</td>
<td>3.02</td>
<td>2.43</td>
<td>2.72</td>
</tr>
<tr>
<td>tbl</td>
<td>2.04</td>
<td>2.10</td>
<td>2.34</td>
<td>2.61</td>
<td>2.80</td>
<td>3.06</td>
</tr>
<tr>
<td>tms</td>
<td>1.99</td>
<td>2.42</td>
<td>1.86</td>
<td>2.21</td>
<td>2.57</td>
<td>2.37</td>
</tr>
<tr>
<td>dfs</td>
<td>2.02</td>
<td>1.92</td>
<td>2.29</td>
<td>2.89</td>
<td>2.53</td>
<td>2.72</td>
</tr>
</tbody>
</table>

#### Across portfolios

#### Market portfolio

<table>
<thead>
<tr>
<th>Predictor</th>
<th>hist avg</th>
<th>no brk</th>
<th>ts</th>
<th>hist avg</th>
<th>no brk</th>
<th>ts</th>
</tr>
</thead>
<tbody>
<tr>
<td>dp</td>
<td>1.59</td>
<td>1.92</td>
<td>2.03</td>
<td>2.50</td>
<td>2.46</td>
<td>2.63</td>
</tr>
<tr>
<td>tbl</td>
<td>1.85</td>
<td>2.02</td>
<td>1.69</td>
<td>2.42</td>
<td>2.64</td>
<td>2.18</td>
</tr>
<tr>
<td>tms</td>
<td>2.12</td>
<td>1.62</td>
<td>1.84</td>
<td>2.59</td>
<td>2.20</td>
<td>2.41</td>
</tr>
<tr>
<td>dfs</td>
<td>1.71</td>
<td>1.47</td>
<td>1.89</td>
<td>2.02</td>
<td>1.59</td>
<td>2.21</td>
</tr>
</tbody>
</table>

**Table 4: Utility gains.** The top panel of this table reports the out-of-sample utility gain for a mean-variance investor with a risk aversion of three who at each period allocates wealth between a risk-free asset (T-bills) and an optimal risky portfolio that is constructed from the 30 industry portfolios. We report the utility gain measured relative to each of the three benchmark models, namely, the prevailing mean (hist avg), the panel model with no breaks (no brk), and the time series break model (ts). The utility gain is computed first using the full sample and second using only the observations that fall within 24 months of a break being detected without counting any observation twice. Results are presented for the four predictors we consider: the dividend-price ratio (dp), the T-bill rate (tbl), the term spread (tms), and the default spread (dfs). The reported values are expressed as annualised percentages. The bottom panel reports the utility gain for a mean-variance investor with a risk aversion of three who allocates his wealth between T-bills and the market portfolio.
Table 5: Dividend growth parameter estimates

<table>
<thead>
<tr>
<th>Regime</th>
<th>Break dates</th>
<th>Intercept Mean</th>
<th>Intercept s.d.</th>
<th>AR(1) Mean</th>
<th>AR(1) s.d.</th>
<th>Dp slope Mean</th>
<th>Dp slope s.d.</th>
<th>Volatility Mean</th>
<th>Volatility s.d.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Feb 1931</td>
<td>-0.038 (0.005)</td>
<td>0.135 (0.039)</td>
<td>-0.082 (0.023)</td>
<td>0.201 (0.003)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>May 1933</td>
<td>0.030 (0.004)</td>
<td>-0.193 (0.031)</td>
<td>0.057 (0.024)</td>
<td>0.156 (0.003)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Aug 1939</td>
<td>-0.127 (0.016)</td>
<td>0.188 (0.043)</td>
<td>-0.034 (0.066)</td>
<td>0.216 (0.010)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Mar 1945</td>
<td>0.021 (0.002)</td>
<td>0.308 (0.022)</td>
<td>-0.029 (0.018)</td>
<td>0.113 (0.001)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Oct 1968</td>
<td>0.019 (0.004)</td>
<td>0.515 (0.032)</td>
<td>-0.156 (0.022)</td>
<td>0.163 (0.003)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Jan 1987</td>
<td>0.027 (0.002)</td>
<td>0.215 (0.029)</td>
<td>-0.267 (0.014)</td>
<td>0.147 (0.002)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Dec 1998</td>
<td>0.002 (0.005)</td>
<td>0.126 (0.045)</td>
<td>-0.305 (0.021)</td>
<td>0.228 (0.004)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Sep 2007</td>
<td>0.072 (0.004)</td>
<td>0.430 (0.041)</td>
<td>-0.169 (0.019)</td>
<td>0.204 (0.003)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>May 2009</td>
<td>0.019 (0.009)</td>
<td>0.307 (0.069)</td>
<td>-0.205 (0.022)</td>
<td>0.327 (0.006)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>0.043 (0.010)</td>
<td>0.834 (0.112)</td>
<td>-0.1872 (0.019)</td>
<td>0.545 (0.007)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5: Dividend growth parameter estimates. This table displays the posterior mean and standard deviation (s.d.) of the intercept, the slope on the AR(1) term and the slope on the lagged dividend-price ratio (Dp slope) obtained from the heterogeneous panel break model in each regime it identifies. The reported values are value-weighted averages across the parameter estimates on the 30 industry portfolios. We also report the mean and standard deviation of the volatility parameter. The posterior modes of the identified break dates are also reported.
Table 6: Sensitivity of results to priors

<table>
<thead>
<tr>
<th>Hyp. value</th>
<th>$K$</th>
<th>DM</th>
<th>CW</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>t $&lt;-1.64$</td>
<td>$-1.64 &lt; t &lt; 0$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No-break panel</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$c=120$</td>
<td>12</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>$c=480$</td>
<td>8</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>$b=1$</td>
<td>10</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>$b=10$</td>
<td>10</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>$\sigma_\eta = 0.02$</td>
<td>10</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>$\sigma_\eta = 0.06$</td>
<td>10</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>$\sigma_\eta = 100$</td>
<td>16</td>
<td>14(\dagger)</td>
<td>8</td>
</tr>
<tr>
<td>$\sigma_\mu = 10%$</td>
<td>10</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Time series breakpoint</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$c=120$</td>
<td>0</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>$c=480$</td>
<td>0</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>$b=1$</td>
<td>0</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>$b=10$</td>
<td>0</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>$\sigma_\eta = 0.02$</td>
<td>0</td>
<td>5</td>
<td>6(\dagger)</td>
</tr>
<tr>
<td>$\sigma_\eta = 0.06$</td>
<td>0</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>$\sigma_\eta = 100$</td>
<td>15(\dagger)</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>$\sigma_\mu = 10%$</td>
<td>0</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>Historical average</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$c=120$</td>
<td>0</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>$c=480$</td>
<td>0</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>$b=1$</td>
<td>0</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>$b=10$</td>
<td>0</td>
<td>5</td>
<td>6(\dagger)</td>
</tr>
<tr>
<td>$\sigma_\eta = 0.02$</td>
<td>0</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>$\sigma_\eta = 0.06$</td>
<td>0</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>$\sigma_\eta = 100$</td>
<td>20(\dagger)</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>$\sigma_\mu = 10%$</td>
<td>1</td>
<td>3</td>
<td>5</td>
</tr>
</tbody>
</table>

Table 6: Sensitivity of the results to priors. This table reports results when forecasting excess returns on the 31 portfolios (including the market) using the heterogeneous panel break model and the dividend-price ratio as the predictive variable. We adjust one hyperparameter at a time thus the hyperparameter value displayed in the table is used alongside all the remaining values detailed in Section Appendix C.1. The three panels display the statistical significance of outperformance or underperformance of our model relative to the three benchmarks we consider. Significance is evaluated using the Diebold-Mariano test (DM) and the procedure of Clark and West (2007) (CW). For each procedure the table displays the number of cases (countries) for which our method produces significantly worse, insignificantly worse, insignificantly better, and significantly better forecasts at the 10% level. † indicates the particular bin in which lies the $t$-statistic for EU-wide inflation forecasts. The top panel of the table also displays the posterior mode of the number of breaks $K$ estimated from our heterogeneous panel break model using the full sample.
Figure 1: This Figure displays how an economic agent could have recursively updated his belief that the global financial crisis had caused a permanent shift in the return prediction model (top window) and how he could have used Bayes’ rule to update in real time the slope coefficient on the lagged dividend-price ratio in his return prediction model (bottom window) when forecasting with the heterogeneous panel break model we propose. The green dotted line in the top window graphs the recursively estimated probability of a break being detected by the time series model with breaks (ts), averaged across the 30 industry portfolios. The solid black line (dashed blue line) denotes the recursively updated posterior model probability assigned to 9 (8) breaks by the panel break model (Pbrk). The vertical red line denotes the time at which the posterior mode switches from 8 to 9 breaks.
Figure 2: This Figure displays the prior probability that the R-squared of a predictive regression lies below a certain value $j$, ranging from 0 to 0.01, for different degrees of scepticism regarding predictability. The degree of scepticism is denoted by the prior standard deviation of the normalised slope coefficient $\sigma_\eta$. A value of 0 denotes a dogmatic prior, a value of infinity denotes a diffuse prior, and intermediate values denote scepticism about the existence of predictability.
Figure 3: This Figure displays the posterior model probabilities (top panel) and the posterior break dates (bottom panel) when regressing the industry portfolios on the lagged dividend-price ratio using the model with pooled breaks and unit-specific parameters and the full sample. The x-axis year labels in the bottom panel mark January of the corresponding years.
Figure 4: This Figure graphs the slope coefficient (top panel) and error-term standard deviation (bottom panel) for the market portfolio over the out-of-sample period. The estimates for the market portfolio are computed as weighted averages of the slopes and volatilities of the 30 industry portfolios recursively estimated from the heterogeneous panel model with pooled breaks that we develop (Pbrk), the same model without breaks (No brk) and the time series break model (ts). The predictive variable is the dividend-price ratio. The x-axis year labels mark January of the corresponding years.
Figure 5: This Figure graphs the forecasts of the market portfolio produced by the panel model with (red dashed line) and without breaks (purple solid line) and the historical average (dotted black line) using the dividend-price ratio as the predictive variable. The heterogeneous panel models with and without breaks are denoted by ‘Pbrk’ and ‘No brk’, respectively, while the industry prevailing mean is denoted by ‘Hist avg’. The labels of the x-axis mark January of the corresponding year.
Figure 6: This Figure displays the real-time break detection when regressing the industry portfolios on the lagged dividend-price ratio using the full sample and the model with pooled break dates and unit-specific parameters. The vertical red line denotes the initial estimation period and the 45 degree line (to the right of the vertical line) denotes the date at which a break could first be identified. The empty black circles mark the estimated break dates.
Figure 7: This Figure displays the number of months it took to first detect each of the breaks that onset after the initial estimation period when predicting with each of the four predictors we consider: the dividend-price ratio (dp), the T-bill rate (tbl), the term spread (tms), and the default spread (dfs).
Figure 8: This Figure graphs the cumulative difference in the sum of squared errors for the portfolio in question obtained from the heterogeneous panel break model developed in this article relative to each of the benchmark models. The benchmark models are the heterogeneous panel model with no breaks (‘No brk’), the industry prevailing mean (‘Hist avg’) and the time-series model with breaks (‘ts’) estimated using the algorithm of Chib (1998) applied to each portfolio in turn. The dividend-price ratio is the predictive variable and the portfolio being forecast is detailed in the subcaption of each subfigure. The x-axis labels mark January of the corresponding years.
Figure 9: This Figure displays the out-of-sample R-squared values obtained when comparing the forecasting performance of the heterogeneous panel break model we develop with the benchmark model in question for each of the thirty industry portfolios and the market portfolio using the dividend-price ratio as the predictive variable. The benchmark model for each subfigure is described in the subcaption. The thick black vertical line marks the out-of-sample R-squared value for the market portfolio.
Figure 10: This Figure graphs the cumulative difference in the sum of squared errors for the portfolio in question obtained from the heterogeneous panel break model developed in this article relative to each of the benchmark models as a function of the time since initial detection of each break which onsets over the out-of-sample period. The squared forecast error is computed as a function of the time since the break was initially detected for each of the breaks that onsets over the out-of-sample period, and the average is taken across these breaks. The competing models are the heterogeneous panel model with no breaks (‘No brk’), the industry prevailing mean (‘Hist avg’), and the time-series model with breaks (‘ts’) estimated using the algorithm of Chib (1998) applied to each portfolio in turn. The dividend-price ratio is the predictive variable and the portfolios being forecast are detailed in the subcaption of each subfigure.
Figure 11: This Figure displays the out-of-sample utility gain to a mean-variance investor who allocates his wealth between the portfolio in question and the risk-free rate. Utility gains are reported as annualised percentages obtained when comparing the forecasting performance of the heterogeneous panel break model we develop with the benchmark model in question for each of the thirty industry portfolios and the market portfolio when using the dividend-price ratio as the predictive variable. The benchmark model for each subfigure is described in the subcaption. The thick vertical black line marks the utility gain for the market portfolio.
Figure 12: This Figure compares the posterior break modes identified by our heterogeneous panel break model when excess returns on the industry portfolios are the dependent variable (red triangles) with the breaks identified when dividend growth of the industry portfolios is the dependent variable (blue triangles) across each of the four predictors we consider: the dividend-price ratio (dp), the T-bill rate (tbl), the default spread (dfs), or the term spread (tms). The model includes an intercept, the lagged predictor in question and when dividend growth is the dependent variable the model further includes an AR(1) term.
Figure 13: This Figure displays the posterior break modes estimated from the heterogeneous panel break model when regressing the dependent variable on an intercept term and the lagged dividend-price ratio. The dependent variable is either the excess returns on the 30 industry portfolios (blue diamonds), the excess returns on the $5 \times 5$ portfolios sorted on size and value (red diamonds) or the excess returns on the $5 \times 5$ portfolios sorted on size and momentum (green circles).
Figure 14: This Figure displays the posterior break dates (top panel) and the real-time break detection (bottom panel) when regressing the industry portfolios on the lagged treasury-bill rate using monthly data and the model with pooled break dates and unit-specific parameters. In the bottom panel the empty black circles denote the estimated break dates and the vertical red line marks the initial estimation period of ten years. The 45 degree line (to the right of the vertical red line) marks the time at which a break could first be detected.
Figure 15: This Figure graphs the cumulative difference in sum squared errors of the market portfolio obtained from the panel break model relative to the three benchmark models. The benchmark models are the heterogeneous panel model with no breaks (‘No brk’), the industry prevailing mean (‘Hist avg’) and the time-series break model with breaks (‘ts’) applied to each portfolio in turn. The predictive variable in question is detailed in the subcaption of each subfigure.