Noncommon Breaks

Simon C. Smith

USC Dornsife INET, Department of Economics, USC

Draft: December 1 2017

Abstract

This paper develops a new Bayesian approach to estimate noncommon structural breaks in panel regression models. Any subset of the cross-section may be hit at different times within a break window. We provide a formal test for noncommon breaks and whether any noncommonality is driven primarily by series not being hit or series being hit with delays. Break-specific pervasion and diffusion parameters are learned from the cross-section and indicate which breaks drive noncommonality. Break pervasiveness gives rise to cross-sectional heterogeneity since only those series hit by a break enter the new regime. In an empirical application to forecasting international stock returns from dividend yields, the method delivers more accurate forecasts than a range of popular benchmark models that are both statistically significant and economically meaningful.

Keywords: Panel data, Structural breaks, Bayesian analysis, Stock return predictability

JEL classifications: G10, C11, C15

Email address: `simonsmi@usc.edu. Mailing address: Institute for New Economic Thinking, Department of Economics, USC, 3620 South Vermont Ave., CA, 90089-0253, USA.` (Simon C. Smith)

Special thanks must go to Allan Timmermann. Also the suggestions of Torben Andersen, Tomohiro Ando, Michael Brennan, Fabio Canova, Alexander Chudik, Giancarlo Corsetti, Frank Diebold, Jean-Marie Dufour, Wayne Ferson, Joao Gomes, Gordon Hanson, Cheng Hsiao, Arthur Korteweg, Peter Koudijs, Dennis Kristensen, Oliver Linton, Roger Moon, Serena Ng, Andrew Patton, Vincenzo Quadrini, Geert Ridder, Roberto Rigobon, Martin Schneider, Ron Smith, Ross Valkanov, Jonathan Wright, Takashi Yamagata, Eric Zivot and participants at USC, NBER-NSF SBIES 2018 and NBER-NSF SI EFFE 2018 have been helpful. Any remaining errors are my own.

1. Introduction

Many macroeconomic and financial time series are subject to structural breaks (Stock and Watson 1996; Pástor and Stambaugh 2001; Chib and Kang 2012; Rossi 2013). Breaks are a useful tool to glean a range of economic insights such as the propagation of macroeconomic shocks across countries or regions in the economy (Hamilton and Owyang 2012); lead-lag predictability that arises from market frictions typically captured through a news diffusion model (Hong et al. 2007; Rapach et al. 2013); the speed of technological advances through the economy (Kapur 1995); and capturing short periods or 'pockets' of return predictability that are interspersed with long periods of little or no predictability (Farmer et al. 2018).

Existing methods for detecting breaks in panels typically assume breaks are common, that is, breaks hit every series in the cross-section at the same time (Bai 2010; Baltagi et al. 2016; Smith and Timmermann 2017b; Smith et al. 2017).[1] This restrictive assumption may not reveal the true underlying break dynamics thereby limiting the economic value of the common break methodology. Applying a time series break method separately to each series has low power and typically overlooks all but the largest breaks (Smith and Timmermann 2017b).[2]

Rather than specifying breaks as a single break date that is common to all series, one can specify a break window that spans multiple periods. Within the break window, any subset of series may be hit at any time. The dynamics of each break window are independent. As an illustrative example, one might ask whether the oil price shock of the mid-1970s represented a break and if so which countries were hit and when? The methodology developed in this article can answer the question by (i) detecting a break window around the mid-1970s if the oil price shock represents a break and (ii) estimating the break dynamics within the window to reveal which countries were hit and when.

This paper develops a new Bayesian approach to estimate an endogenously determined number of noncommon breaks at unknown times in a panel regression model.

[1] A subset of studies on breaks in panel models or multivariate time series include Bai et al. (1998), De Wachter and Tzavalis (2005), Qu and Perron (2007), Bai and Carrion-I-Silvestre (2009), Kim (2011), Oka and Qu (2011), De Wachter and Tzavalis (2012), Horváth and Hušková (2012), Groen et al. (2013), Kim (2014), Qian and Su (2016), and Baltagi et al. (2017).

[2] A handful of frequentist time series approaches to detect breaks include Andrews (1993), Bai and Perron (1998), Bai and Perron (2003), and Elliott and Müller (2006) while Bayesian approaches include Chib (1998), Wang and Zivot (2000), Pesaran et al. (2006), Koop and Potter (2007), Geweke and Jiang (2011), Giordani and Kohn (2012), and Peluso et al. (2016).

Each break can hit any subset of series at different times within a break window. This allows cross-sectional hetereogeneity in (i) which series are hit, (ii) the times they are hit and (iii) the uncertainty surrounding both. A Bayesian common break approach fails to capture the cross-sectional heterogeneity in all three, while a frequentist non-common break approach would not capture the uncertainty.

While noncommon breaks offer the opportunity to better fit the data, it comes at the expense of increased model complexity. Common breaks are nested within our framework if each break hits every series simultaneously. If our method estimates breaks to be noncommon, it is important to formally test whether the data truly justify noncommon relative to common breaks while accounting for increased model complexity.[3]

To the best of our knowledge, this is the first paper to endogenously estimate and formally test for noncommon breaks in panel models. Using the approach of Chib (1995) we compute the marginal likelihoods for the noncommon and common breaks models and the corresponding Bayes Factor. This Bayes Factor evaluates the strength of evidence in favour of noncommon breaks while integrating over all model parameters and inherently penalising overparameterisation.

Methods for testing whether structural breaks in time series occur smoothly or abruptly are typically used to approximate the nature of shifts in the aggregate (Elliott and Müller 2006; Koop and Potter 2010; Chen and Hong 2012). Such methods, however, are silent on what is driving the occurrence and nature of breaks. Our method reveals which series in the disaggregate panel are hit and at what times. Understanding the underlying dynamics of such breaks is critical to companies and policy-makers. For instance, companies such as Amazon that wish to forecast the future sales performance of various product lines must continuously monitor which product lines are likely to be affected by breaks. Likewise, the Federal Reserve must adjust interest rates in a timely manner to meet its inflation targets. Since the target is the U.S. aggregate, which is a weighted average of inflation in the various U.S. industries, policy-makers must continuously evaluate which, if any, industries have experienced breaks such that their forecasting models can be updated accordingly.

We further test whether any noncommonality is driven primarily by diffusion or a lack of pervasion. We estimate and compute marginal likelihoods for two versions

[3]There exists a long history of testing for both the existence and the nature of structural breaks, particularly in time series models (Andrews 1993; Andrews and Ploberger 1994; Bai and Perron 1998; Andrews 2003; Elliott and Müller 2006; Chen and Hong 2012).

of the model that either restrict breaks to (i) diffuse through the entire cross-section with different delays or (ii) hit a subset of series at the same time.

In addition, break-specific diffusion and pervasion parameters are learned from the cross-section. We can thus glean *which* breaks are driving any noncommonality and whether it is driven primarily by nonpervasion or diffusion. These parameters provide the model with the desirable property that series inform one another regarding pervasion and diffusion. For instance, if 90% of series have been observed to be hit by a break the probability of the remaining 10% of series being hit increases. Similarly, if the majority of series observed thus far have been hit with long delays the probability of the remaining series being hit with long delays is higher.

To illustrate the importance of allowing for noncommon breaks observe the plots in Figure 1. The top left window displays the countries hit by the break around the oil price shock of the mid-1970s. The breaks are estimated from a pooled panel regression of excess stock returns for 17 countries on an intercept and lagged country-level dividend yields.[4] While the common break model imposes that every country is hit (blue crosses), the noncommon breaks model (red triangles) reveals that Hong Kong, Malaysia and Singapore are not. Identifying that the oil price shock did not hit these three countries and that they experienced a longer initial period of parameter stability would have been valuable information for investors. Specifically, the common breaks model (blue line in bottom window) assumes that all countries became considerably less predictable following the oil price shock with their slope coefficient falling from approximately 4 to 0.5. The noncommon breaks model, however, tells a similar story for the 14 countries hit by the break (red line) but also reveals that Hong Kong, Malaysia and Singapore (black line) remain equally predictable after the oil price shock.

A key innovation relative to existing studies is the ability to not only detect which series are hit by each break and at what time, but also the uncertainty surrounding both. Incorporating the cross-sectional heterogeneity of the uncertainty through both the pervasion and diffusion channels into the parameter estimates may be critical to avoid inference being compromised.

With such flexible break dynamics the combinatorial possibilities render infeasible an exhaustive search typically implemented by frequentist approaches such as Bai and Perron (1998).[5] A Bayesian approach is desirable whereby estimation employs

[4] A more detailed description of the data is provided in Section 6.

[5] Allowing any subset to be hit at the same date would involve 2^N possibilities which becomes

a stochastic search algorithm that does not require an exhaustive search, but more frequently visits regions of the posterior parameter space with high density.

The model is estimated using the reversible jump Markov chain Monte Carlo algorithm developed by Green (1995) (see also Carlin and Chib (1995)). Estimation consists of four steps. The first step samples the parameters from their full conditionals using the Gibbs sampler. The second step estimates the timing of each break window in turn using a random-walk Metropolis-Hastings algorithm. The third step estimates within each break window which series are hit and when. The fourth step estimates the number of breaks by attempting with equal probability to either introduce a new break window at a randomly chosen location with break dynamics sampled in the manner described in step three or remove a randomly selected existing break window.

A simulation study in which the underlying data generating process is known is conducted to illustrate the usefulness of the method to estimate and evaluate the evidence in favour of noncommon breaks. The method is able to detect breaks in both level and variance that range from common breaks to breaks that hit only a subset of series at different times. Crucially, it correctly identifies which subset of series are hit by breaks that are not fully pervasive. The heterogeneous common break model overlooks the break altogether which biases the parameter estimates of those series hit by the break in the two regimes it separates. The pooled common break model detects the nonpervasive break but incorrectly forces all series to shift thereby biasing the parameter estimates of all series in the subsequent regime. The more (less) pervasive the break, the more bias is induced by the heterogeneous (pooled) common break model. The noncommon breaks model correctly identifies which series are hit thereby estimating unbiased parameters and allocating them to the appropriate series.

The practical value of the framework is illustrated in an empirical application to the out-of-sample predictability of excess stock returns of 17 countries over the period 1973-2017 using their corresponding dividend yields. The noncommon breaks model detects four breaks over the 45-year sample that correspond to key economic events and reveals which countries are hit and when. The U.S. is always hit first while the remaining countries are typically hit with a one or two month lag supporting the findings of Rapach et al. (2013) that shocks either originate in or first hit the U.S. and are subsequently propagated to other countries giving rise to a leading role for

infeasible even for moderate cross-sectional dimensions N. Allowing series to be hit at different times exacerbates the problem even further.

the U.S. in international stock return predictability.

We compare the out-of-sample predictive accuracy of stock return forecasts from our model to five benchmarks that include the country-level prevailing mean, a time series linear predictive regression applied independently to each country, a pooled constant panel model, and two panel common break models with either pooled or unit-specific parameters. Our model produces more accurate forecasts for the majority of countries that are consistent across the out-of-sample period, statistically significant and economically meaningful for risk averse mean-variance investors who at each time period allocate their wealth between risk-free and risky assets.

The remainder of the paper is set out as follows. Section 2 presents the model and details the prior specifications. Section 3 explains estimation of the model. Section 4 presents the formal test of noncommon breaks. Section 5 presents the simulation study. Section 6 presents an empirical application to international excess stock return predictability. Section 7 presents robustness checks and extensions. Section 8 concludes.

2. Methodology

Before providing the specific details of the methodology, we briefly discuss the benefits of estimating break dynamics within break windows relative to an unconstrained search. From an economic perspective, structural breaks, which typically correspond to major economic events such as the oil price shocks of the 1970s or the global financial crisis, are likely to diffuse through the economy within relatively short periods. From a statistical perspective, searching for break dynamics within break windows is considerably less prohibitive than a completely unconstrained search over all possible break combinations.

In order for the model to be identified, each break window must have at least one series being hit at the first and final time points of the window. If we estimate a series to be hit beyond the end of the break window, the break window is extended (up to an upper bound that is pre-specified by the user). Likewise, if we estimate that the final period of the window contains no series hit, we shorten the window until the final time point of the window contains a series that is hit. Any estimate that involves the first time point of the window containing no series being hit is rejected, but the location of the window can be moved in either direction.

We consider a panel regression model with $i = 1, \ldots, N$ series and $t = 1, \ldots, T$ time periods that is subject to an endogenously estimated number of break windows

K that separate $K+1$ regimes. Parameters are pooled within regimes. Within the kth break window, let $\mathbb{1}_{ik}$ denote an indicator function which equals one if the ith series is hit by the kth break and zero otherwise. Further let $\mathbb{1}_k = (\mathbb{1}_{1k}, \ldots, \mathbb{1}_{Nk})$ and $\mathbf{1} = (\mathbb{1}_1, \ldots, \mathbb{1}_K)$. The kth break hits $N_k = \sum_{i=1}^{N} \mathbb{1}_{ik}$ series, in which $1 \leq N_k \leq N$.[6] Since some series may not be hit by every break, the unit-specific number of breaks, K_i, may vary among series, subject to the restriction $0 \leq K_i \leq K$.

For those series $i \in N_k$ hit by the kth break we endogenously estimate the delay with which they are hit. Let Δ_{ik} denote the delay with which the ith series is hit by the kth break, $\Delta_k = (\Delta_{1k}, \ldots, \Delta_{Nk})$ and $\mathbf{\Delta} = (\Delta_1, \ldots, \Delta_K)$. We therefore have a break window for the kth break which spans the time from the onset of the break through the time at which the final series is hit. Let $\Delta_{k_{max}} = max(\Delta_{1k}, \ldots, \Delta_{Nk})$ denote the length of the kth break window subject to the restriction $\Delta_{k_{max}} \leq \Delta_{max}$ in which Δ_{max} is an upper bound on the break window that is pre-specified by the user. The common break model assumes $\Delta_{max} = 0$ such that every series is hit at the same time while at the other extreme $\Delta_{max} = T$ essentially applies a time series break model (allowing at most one break) separately to each series.

Let $\tau = (\tau_1, \ldots, \tau_{2K})$ denote a breakpoint vector characterising the break windows where τ_{2k-1} denotes the time point that immediately precedes the start of the kth break window and $\tau_{2k} = \tau_{2k-1} + \Delta_{k_{max}}$ the final time point of the kth break window. The duration of the kth regime is denoted l_k. Since some series may not be hit by any breaks let $N_{K\geq 0}$ denote the total number of series that experience at least one break. For $i \in N_{K\geq 0}$ we have a breakpoint vector $\tau_i = (\tau_{1_i}, \ldots, \tau_{K_i})$ and let $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_{N_{K\geq 0}})$.[7] The duration of the k_ith regime is denoted l_{k_i}.

The framework nests (i) the common break pooled panel approach in which every series is hit by a break at the same time (Bai 2010; Baltagi et al. 2016; Smith and Timmermann 2017b), (ii) with a sufficiently diffuse prior on the regime durations, a panel time-varying parameter model in which a (typically small) break occurs every period, (iii) a constant pooled parameter panel model, and (iv) partial break models in which any subset of series are hit by any given break.

[6] We assume $N_0 = N$.

[7] For convenience we assume throughout that $\tau_0 = 0$ and $\tau_{2K+1} = T$, and, for all i, $\tau_{0_i} = 0$ and $\tau_{K_i+1} = T$.

2.1. Modeling cross-sectional dependencies

The majority of panel data sets used in macroeconomics and finance are likely to have cross-sectional dependencies. Ignoring such dependencies can compromise inference (Andrews 2005) and eliminate much of the increased break detection power derived from the cross-section (Kim 2011; Baltagi et al. 2016). A popular method for accounting for cross-sectional dependencies is the common factor approach - see, for example, Pesaran (2006) and Bai (2009) - which includes one or more common factors to absorb the dependencies and is particularly straightforward if the factor is observed.

Our framework includes observed common factors to account for cross-sectional dependencies and assumes errors are cross-sectionally independent thereafter. In Section 6, we conduct the formal cross-sectional (CD) test of Pesaran (2004) to evaluate the success of including two observed common factors to remove any cross-sectional dependencies in the empirical application.

Nonetheless the focus of this article is the development of a new methodology that generates more accurate forecasts than a range of popular methods when forecasting in the presence of structural breaks. Since we are not performing inference on the parameter estimates any cross-sectional dependence that remains is likely to be of little concern. The ultimate test of our model is whether it generates significantly more accurate international stock return forecasts compared with a range of benchmark models that include the consistently estimated linear predictive regression.

2.2. Model with pooled breaks and parameters

Let $\mathbf{y}$ denote a $(N \times T)$ matrix of observations on the dependent variable and $\mathbf{X}$ denote a $(N \times T \times \kappa)$ three-dimensional array of observations on the κ regressors. For regimes $k = 1, \ldots, K+1$ the panel regression with pooled parameters and noncommon breaks takes the form

$$y_{it} = X_{it}'\beta_k + \epsilon_{it}, \qquad i = 1, \ldots, N_{k-1}, \qquad t = \tau_{k_i-1} + 1, \ldots, \tau_{k_i} \tag{1}$$

in which we assume Gaussian errors $\epsilon_{it} \sim N(0, \sigma_k^2)$.[8] Let $\beta_k = (\beta_{1,k}, \dots, \beta_{\kappa,k})$, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_{K+1})$ and $\sigma^2 = (\sigma_1^2, \dots, \sigma_{K+1}^2)$. Given the specification of the regression model and the assumption of Gaussian error-term variances the likelihood is

$$p(\boldsymbol{y} \mid \boldsymbol{X}, \boldsymbol{\beta}, \sigma^2, \boldsymbol{\tau}) = \prod_{k=1}^{K+1} \prod_{i=1}^{N_{k-1}} (2\pi\sigma_k^2)^{-l_{k_i}/2} exp \left[-\frac{1}{2} \sum_{k=1}^{K+1} \sum_{i=1}^{N_{k-1}} \sum_{t=\tau_{k_i-1}+1}^{\tau_{k_i}} \frac{(y_{it} - X_{it}'\beta_k)^2}{\sigma_k^2} \right]. \tag{2}$$

2.3. Cross-sectional heterogeneity from break dynamics

The parameters of panel models are typically specified as either unit-specific, estimates of which will be unbiased but imprecise, or pooled across the entire cross-section which will increase the precision at the expense of inducing bias. A compromise is to specify parameters that are pooled within, but differ across, groups (Bonhomme and Manresa 2015). We infer cross-sectional heterogeneity from break pervasiveness.

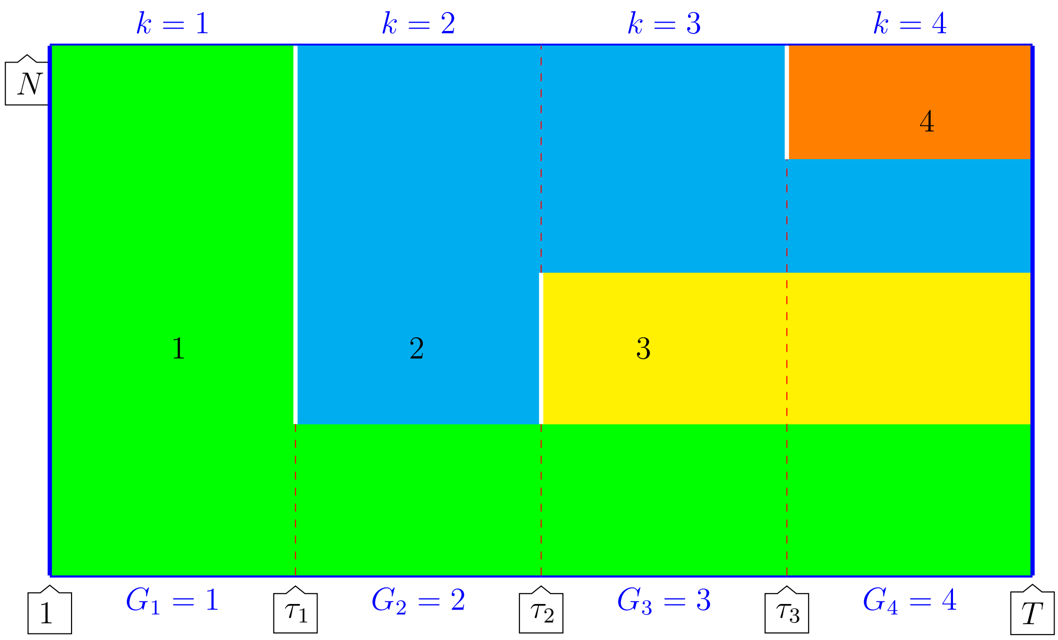
Our framework pools parameters within regimes to increase the precision of the estimate but allows cross-sectional heterogeneity because some series may not be hit by breaks and thus series may be in different regimes at a given time. Specifically, the cross-sectional heterogeneity is inferred directly from the break dynamics: those series hit by a break enter the new regime while the remaining series stay in the current regime. The Figure embedded in the text illustrates this point. The x-axis denotes time and the y-axis denotes the N series. In regime one all series belong to regime one (shaded in green). At time $t = \tau_1$ a break hits only the subset of series that correspond to the vertical solid white line and they enter the new regime (shaded in blue). The remaining series not hit correspond to the vertical red dashed line and stay in regime one. We now have some series in regime one and some in regime two thereby introducing cross-sectional heterogeneity. At time $t = \tau_2$ another break hits another subset of series which subsequently enter regime three (yellow). The series not hit remain in their current regimes. By the end of the sample we have experienced three nonpervasive breaks and have four different regimes with some series in each.

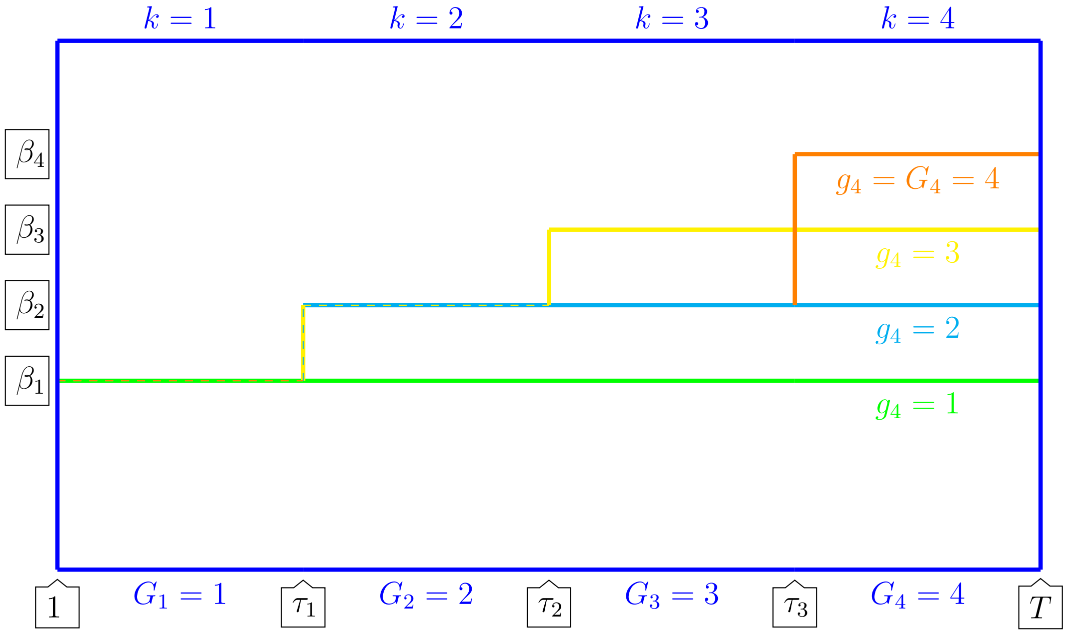
Furthermore, the parameters of those series not hit by a break remain constant thereafter and are estimated using the appropriate pre- and post-break data thereby increasing the precision of the estimate. For instance, after the third break the pa-

[8]Allowing the first of the κ regressors to be a unit vector results in the first element of β_k being an estimate of the pooled intercept in the kth regime. Observed common factors and their loadings are suppressed here for ease of exposition.

rameters of regime one are estimated using the entire area shaded green in the Figure.

There are four groups in total whereby the slope coefficient β_1 is pooled across the green area, β_2 across the blue area, β_3 the yellow, and β_4 the orange. The same grouping and pooling structure applies to all regression coefficients. This gives rise to four different possible paths of regression coefficients for each of the N series to follow. The bottom panel of the embedded Figure illustrates the evolution of the slope coefficient for the four groups. The green horizontal solid line graphs the evolution of all series within group one. These series are not hit by any of the breaks and thus have a constant slope coefficient β_1. The blue line illustrates that series within group two have a slope coefficient equal to β_1 until time $t = \tau_1$ after which they shift upwards to β_2 and remain constant hereafter. The series in groups three and four are graphed by the yellow and orange lines.





2.4. Prior distributions

Bayesian methodology combines information in the data captured through the likelihood function with prior information. The prior distributions are now specified.

2.4.1. Prior on regime durations

We follow Koop and Potter (2007) and Smith and Timmermann (2017b) by assuming regime durations have a Poisson distribution

$$p(l_k \mid \lambda_k) = \frac{\lambda_k^{l_k}}{l_k!} exp^{-\lambda_k}, \qquad k = 1, \dots, K+1 \tag{3}$$

in which the hyperparameter λ_k, which denotes the expected duration of the kth regime, has a conjugate Gamma prior distribution

$$p(\lambda_k) = \frac{d^c}{\Gamma(c)} \lambda_k^{c-1} exp^{-\lambda_k d}, \qquad k = 1 \dots, K+1. \tag{4}$$

Smith and Timmermann (2017b) show that λ can be marginalized obtaining

$$p(l) = \prod_{k=1}^{K+1} \frac{1}{l_k!} \frac{\Gamma(c + l_k)}{(d+1)^{c+l_k}} \frac{d^c}{\Gamma(c)}. \tag{5}$$

2.4.2. Prior on pervasiveness of breaks

The pervasiveness of the kth break is captured by the parameter p_k. The higher the value of p_k, which is subject to the condition $0 \leq p_k \leq 1$, the more pervasive the kth break.

For series $i = 1, \ldots, N$ and breaks $1, \ldots, K$, we specify a Bernoulli prior distribution on the indicator $\mathbb{1}_{ik}$ such that $p(\mathbb{1}_{ik} \mid p_k) = p_k$ if $\mathbb{1}_{ik} = 1$ and $1 - p_k$ if $\mathbb{1}_{ik} = 0$. If we have observed a large number of series hit by the kth break we revise upwards our estimate of p_k which in turn implies that any yet unobserved series have an increased probability of being hit. The converse also holds. The probability of each series being hit by the kth break is characterised by p_k which has a conjugate Beta prior distribution

$$p(p_k) = p_k^{g-1}(1-p_k)^{h-1}\frac{\Gamma(g+h)}{\Gamma(g)\Gamma(h)}, \qquad k = 1, \ldots, K. \tag{6}$$

Multiplying and dividing by $\frac{\Gamma(g+h+N)}{\Gamma(g+N_k)\Gamma(h+N-N_k)}$ and collecting and rearranging terms we can marginalize p

$$p(\mathbb{1}) = \prod_{k=1}^{K}\int\left(\prod_{i=1}^{N}p(\mathbb{1}_{ik} \mid p_k)\right)p(p_k)dp_k = \prod_{k=1}^{K}\frac{\Gamma(g+h)}{\Gamma(g)\Gamma(h)}\frac{\Gamma(g+N_k)\Gamma(h+N-N_k)}{\Gamma(g+h+N)}. \tag{7}$$

2.4.3. Prior on diffuseness of breaks

Our framework allows series to be hit by breaks with different delays. Diffusion dynamics are break-specific and thus one break may diffuse faster through the cross-section than another. We specify one prior over the regime durations and another over the break intervals to enable the value of the parameter δ_k to be learned from the cross-section. The more series hit by the kth break with longer delays the more we revise upwards our estimate of δ_k which, in turn, increases the probability of any yet unobserved series being hit with longer delays.

For each series hit by the kth break we specify a Poisson prior on its break delay

$$p(\Delta_{ik} \mid \delta_k, \mathbb{1}_{ik}) = \frac{\delta_k^{\Delta_{ik}}}{\Delta_{ik}!}exp(-\delta_k), \qquad k = 1, \ldots, K, \qquad i = 1, \ldots, N_k \tag{8}$$

in which the hyperparameter δ_k, which denotes the average expected lag with which the N_k series are hit by the kth break, has a conjugate Gamma prior distribution

$$p(\delta_k) = \frac{f^e}{\Gamma(e)}\delta_k^{e-1} exp(-f\delta_k) \qquad k = 1, \dots, K. \tag{9}$$

For the kth break let $\Delta_{N_k} = \sum_{i=1}^{N_k} \Delta_{ik}$. Multiplying and dividing by $\frac{(f+N_k)^{e+\Delta_{N_k}}}{\Gamma(e+\Delta_{N_k})}$ and collecting and rearranging terms we can marginalize δ

$$p(\Delta) = \prod_{k=1}^{K}\prod_{i=1}^{N_k}\int p(\Delta_{ik} \mid \delta_k)p(\delta_k)d\delta_k = \prod_{k=1}^{K}\left(\frac{\Gamma(e+\Delta_{N_k})}{(f+N_k)^{e+\Delta_{N_k}}}\frac{f^e}{\Gamma(e)}\prod_{i=1}^{N_k}\left(\frac{1}{\Delta_{ik}!}\right)\right). \tag{10}$$

Finally let $p(\boldsymbol{\tau}) = p(l)p(\mathbb{1})p(\Delta)$.

2.4.4. Priors on regression and variance parameters

To preserve conjugacy, which is critical to the mixing and computational efficiency of estimation, the error-term variances have an inverse gamma prior distribution

$$p(\sigma_k^2) = \frac{b^a}{\Gamma(a)}\sigma_k^{2^{-(a+1)}} exp\left(-\frac{b}{\sigma_k^2}\right), \qquad k = 1, \dots, K+1 \tag{11}$$

and the intercept and slope coefficients have a Gaussian prior distribution conditional on σ_k^2

$$\begin{aligned} p(\beta_k \mid \sigma_k^2) &= 2\pi^{-\kappa/2}(\sigma_k^2)^{-\kappa/2} \mid V_\beta \mid^{-1/2} exp\left(-\frac{1}{2\sigma_k^2}\beta_k' V_\beta^{-1}\beta_k\right), \qquad k = 1, \dots, K+1 \\ V_\beta &= \sigma_\beta^2 I_\kappa \end{aligned} \tag{12}$$

in which a and b characterise the a priori expected error-term variance and σ_b^2 characterises the prior variance of β.

2.5. Posterior distribution

Inference is performed on the posterior distribution which combines prior information supplied by the user with information in the data transmitted through the likelihood function. Letting $l_{N_{k-1}} = \sum_{i=1}^{N_{k-1}} l_{k_i}$ the posterior is

$$p(\boldsymbol{\beta}, \sigma^2 \mid \boldsymbol{y}, \boldsymbol{X}, \boldsymbol{\tau}) = \left(\prod_{k=1}^{K+1} (2\pi)^{-\left(\kappa/2 + l_{N_{k-1}}/2\right)} (\sigma_k^2)^{-\left(a+1+\kappa/2+l_{N_{k-1}}/2\right)} \frac{b^a}{\Gamma(a)} \mid V_\beta \mid^{-1/2} \right)$$
$$\times\, exp \left[\sum_{k=1}^{K+1} -\frac{1}{2\sigma_k^2} \left(2b + \beta_k' V_\beta^{-1} \beta_k + \sum_{i=1}^{N_{k-1}} \sum_{t=\tau_{k_i-1}+1}^{\tau_{k_i}} (y_{it} - X_{it}'\beta_k)^2 \right) \right] \tag{13}$$

Specifying conjugate priors on the regression coefficients enables them to be marginalized enhancing the computational efficiency of the method which is crucial when implementing the algorithm on large panels. Appendix A details how $\boldsymbol{\beta}$ and σ^2 can be marginalized from the posterior in equation (13) to obtain

$$p(\boldsymbol{y} \mid \boldsymbol{X}, \boldsymbol{\tau}) = 2\pi^{-TN/2} \prod_{k=1}^{K+1} \frac{b^a}{\Gamma(a)} \frac{\Gamma(\tilde{a}_k)}{\tilde{b}_k^{\tilde{a}_k}} \frac{\mid \Sigma_k \mid^{1/2}}{\mid V_\beta \mid^{1/2}} \tag{14}$$

in which, for regimes $k = 1, \ldots K + 1$

$$\begin{aligned}
\Sigma_k^{-1} &= V_\beta^{-1} + \sum_{i=1}^{N_{k-1}} \sum_{t=\tau_{k_i-1}+1}^{\tau_{k_i}} X_{it} X_{it}' \\
\mu_k &= \Sigma_k \times \sum_{i=1}^{N_{k-1}} \sum_{t=\tau_{k_i-1}+1}^{\tau_{k_i}} X_{it} y_{it} \\
\tilde{a}_k &= a + \frac{l_{N_{k-1}}}{2} \\
\tilde{b}_k &= \frac{1}{2} \left(2b - \mu_k' \Sigma_k^{-1} \mu_k + \sum_{i=1}^{N_{k-1}} \sum_{t=\tau_{k_i-1}+1}^{\tau_{k_i}} y_{it}^2 \right).
\end{aligned} \tag{15}$$

3. Estimating the Model

Estimating the model using Markov chain Monte Carlo methods consists of four stages. First, the regime-specific parameters are sampled from their full conditional distributions using the Gibbs step. Second, the timing of each break window is estimated by perturbing it using a random-walk Metropolis-Hastings algorithm. Third, the break dynamics within each window are estimated. Fourth, to estimate the number of breaks we allow for a new break window to be introduced (birth move) or an existing one to be removed (death move). This estimation step jointly estimates the number, timing and dynamics of breaks.

3.1. Estimating the parameters

In turn, we draw each of the parameters from their full conditional distributions. For regimes $k = 1, \ldots, K+1$ we sample

$$\begin{aligned} p(\lambda_k \mid \cdots) &\sim Gamma(c + l_k, d + 1) \\ p(\sigma_k^2 \mid \cdots) &\sim IG\left(\tilde{a}_k, \tilde{b}_k\right) \\ p(\beta_k \mid \cdots) &\sim N\left(\mu_k, \Sigma_k \sigma_k^2\right) \end{aligned} \tag{16}$$

in which the hyperparameters are computed using equation (15). For the $k = 1, \ldots, K$ break windows we sample

$$\begin{aligned} p(\delta_k \mid \cdots) &\sim Gamma\left(e + \Delta_{N_k}, f + N_k\right) \\ p(p_k \mid \cdots) &\sim Beta\left(g + N_k, h + N - N_k\right). \end{aligned} \tag{17}$$

3.2. Estimating the location of break windows

To estimate the location of break windows we perturb the location of the kth break window to obtain a proposed window location. Specifically, the timing of the kth break window is perturbed by u which is sampled from the discrete uniform interval $u \sim U[-s, s]$ such that $\tau_{2k^*-1} = \tau_{2k-1} + u$ and $\tau_{2k^*} = \tau_{2k} + u$, and, for all i hit by the kth break, $\tau_{k_i^*} = \tau_{k_i} + u$ while the dynamics of the break window remain unchanged. If $u = 0$ the move is immediately rejected. The length of regimes k and $k+1$ are computed as $l_{k^*} = l_k + u$ and $l_{k^*+1} = l_{k+1} - u$, and, for all i, $l_{k_i^*} = l_{k_i} + u$ and $l_{k_i^*+1} = l_{k_i+1} - u$. Using these proposals we construct l^*, τ^* and, for all i, l_i^* and τ_i^*. Let l_{min} denote a minimum regime duration pre-specified by the user.[9] If l_{k^*} or l_{k^*+1} is less than l_{min} the proposal is rejected. Otherwise, we compute $\Sigma_{k^*}^{-1}$, μ_{k^*}, $\tilde{a}_{k^*}$, and $\tilde{b}_{k^*}$ from equation (15) for the $k = 1, \ldots, K+1$ regimes. We accept the proposal with probability $min(1, \alpha)$

$$\alpha = \left(\prod_{k=1}^{K+1} \frac{\tilde{b}_k^{\tilde{a}_k}}{\tilde{b}_{k^*}^{\tilde{a}_{k^*}}} \frac{\Gamma(\tilde{a}_{k^*})}{\Gamma(\tilde{a}_k)} \frac{\mid \Sigma_{k^*} \mid^{1/2}}{\mid \Sigma_k \mid^{1/2}} \frac{l_k!}{l_{k^*}!} \frac{\Gamma(c + l_{k^*})}{\Gamma(c + l_k)} \right) \tag{18}$$

[9] In the simulation study and empirical application that follow $l_{min} = 1$.

in which terms cancel because of constants and, for all i, $\sum_{k=1}^{K+1} l_{k_i} = \sum_{k=1}^{K+1} l_{k_i^*}$. If the proposal is accepted, we substitute all proposal for their current values. Otherwise the proposal is discarded and the algorithm continues to the next step.

3.3. Estimating the break window dynamics

For each series $i = 1, \ldots, N$ in turn, we sample whether the ith series is hit by the k^*th break from a Bernoulli distribution with equal probability, $\mathbb{1}_{ik^*} \sim Bern(0.5)$ and construct $\mathbb{1}_{k^*}$. If the ith series is hit ($\mathbb{1}_{k^*}$=1), we sample the delay with which it is hit uniformly from the discrete interval $\Delta_{ik^*} \sim U[0, \Delta_{max}]$. We compute K_i^*. If $\mathbb{1}_{k^*} = \mathbb{1}_k$ and $\Delta_{ik^*} = \Delta_{ik}$, the move is rejected. If the sampled value of Δ_{ik^*} violates the assumption that at least one series is hit immediately by the break we set τ_{2k^*-1} equal to the date at which a series is first hit by the k^*th break and thus re-compute Δ_{k^*}, $\Delta_{k^*_{max}}$ and l_{k^*} accordingly. Likewise if the ith series is sampled to be hit by the k^*th break beyond the end of the current break interval $\tau_{k_i^*-1} > \tau_{2k}$ we extend the break window by setting τ_{2k^*} equal to this value. We subsequently compute $\Delta_{k^*_{max}}$ and l_{k^*+1}.

If $\Delta_{k^*_{max}} > \Delta_{k_{max}}$ the length of the k^*th break window has increased and the subsequent regime duration l_{k^*+1} has shortened, but if $\Delta_{k^*_{max}} < \Delta_{k_{max}}$ the length of the k^*th break window has decreased and the subsequent regime duration l_{k^*+1} has extended. This is crucial in allowing the length of each break window to vary (within the upper bound prespecified by the user).

We construct l^*, τ^*, l_i^*, τ_i^*, N_{k^*}, $\Delta_{N_{k^*}}$, ${\Sigma^*}^{-1}$, μ^*, $\tilde{a}^*$, and $\tilde{b}^*$. The proposal is accepted with probability $min(1, \alpha)$ whereby

$$\alpha = \left(\prod_{k=1}^{K+1} \frac{\Gamma(\tilde{a}_k^*)}{\Gamma(\tilde{a}_k)} \frac{\tilde{b}_k^{\tilde{a}_k}}{\tilde{b}_k^{*\tilde{a}_k^*}} \frac{| \Sigma_k^* |^{1/2}}{| \Sigma_k |^{1/2}} \right) \left(\prod_{i=1}^{N} \frac{\Delta_{ik}!}{\Delta_{ik^*}!} \right) \frac{(d+1)^{l_k}}{(d+1)^{l_{k^*}}} \frac{l_k!}{l_{k^*}!} \frac{\Gamma(c + l_{k^*})}{\Gamma(c + l_k)} \frac{(d+1)^{l_{k+1}}}{(d+1)^{l_{k^*+1}}}$$
$$\times \frac{l_{k+1}!}{l_{k^*+1}!} \frac{\Gamma(c + l_{k^*+1})}{\Gamma(c + l_{k+1})} \frac{\Gamma(e + \Delta_{N_{k^*}})}{\Gamma(e + \Delta_{N_k})} \frac{(f + N_k)^{e + \Delta_{N_k}}}{(f + N_{k^*})^{e + \Delta_{N_{k^*}}}} \frac{\Gamma(g + N_{k^*})\Gamma(h + N - N_{k^*})}{\Gamma(g + N_k)\Gamma(h + N - N_k)}. \quad (19)$$

If the proposal is rejected we discard it, otherwise we substitute all proposals for their original values.

3.4. Estimating the number of break windows

Our framework estimates the number of break windows endogenously. On each sweep of the Markov chain Monte Carlo run with equal probability the algorithm attempts to either introduce a new break window at a randomly chosen location (birth move) or remove a randomly chosen existing break window (death move).

3.4.1. Birth move

A birth move is entered with probability $b_K = 0.5$ and attempts to add a break window k^* thereby increasing K to $K^* = K + 1$.[10] For series $i = 1, \ldots, N$ we propose with equal probability a value of one or zero indicating which of the N series are hit by the k^*th break, i.e. $\mathbb{1}_{ik^*} \sim Bern(0.5)$. For those series hit by the break, we sample the delay with which they are hit uniformly from the discrete interval $\Delta_{ik^*} \sim U[0, \Delta_{max}]$. We compute N_{k^*} and $\Delta_{N_{k^*}}$. If no series is hit by the break without delay the move is rejected. Otherwise we compute $\Delta_{k^*_{max}}$, uniformly sample from the discrete time series the final time point before the proposed break interval begins $\tau_{2k^*-1} \sim U[1, T]$, and compute $\tau_{2k^*} = \tau_{2k^*-1} + \Delta_{k^*_{max}}$.

We are proposing to split the current regime k into two new shorter regimes k^* and $k^* + 1$ separated by the k^*th break interval of length $\Delta_{k^*_{max}}$. The k^*th break interval consists of the observations $\tau_{2k^*-1} + 1, \ldots, \tau_{2k^*}$. Let $l_{k^*} = \tau_{2k^*-1} - \tau_{2(k^*-1)}$ since regime k^* contains the observations $\tau_{2(k^*-1)} + 1, \ldots, \tau_{2k^*-1}$ and $l_{k^*+1} = \tau_{2(k^*+1)-1} - \tau_{2k^*}$ since regime $k^* + 1$ contains the observations $\tau_{2k^*} + 1, \ldots, \tau_{2(k^*+1)-1}$. The move is rejected if the proposed break interval overlaps with an existing break interval ($\tau_{2k^*-1} \leq \tau_{k-1} + l_{min}$ or $\tau_{2k^*} \geq \tau_k - l_{min}$) or if either proposed regime duration is shorter than the prespecified minimum (l_{k^*} or $l_{k^*+1} < l_{min}$).

We further compute τ^* and l^*, and $\Sigma^{*^{-1}}$, μ^*, $\tilde{a}^*$, and $\tilde{b}^*$ from equation (15) and, for all i, we construct l_i^* and τ_i^*. The move is accepted with probability $min(1, \alpha)$

$$\begin{aligned}
\alpha = &\frac{b^a}{\Gamma(a) \mid V_\beta \mid^{1/2}} \left(\prod_{k=1}^{K^*+1} \frac{\Gamma(\tilde{a}_k^*) \mid \Sigma_k^* \mid^{1/2}}{\tilde{b}_k^{*\tilde{a}^*_k}} \right) \left(\prod_{k=1}^{K+1} \frac{\tilde{b}_k^{\tilde{a}_k}}{\Gamma(\tilde{a}_k) \mid \Sigma_k \mid^{1/2}} \right) \\
&\times \frac{2^N T (\Delta_{max} + 1)^{N_{k^*}} l_k!}{(K+1) l_{k^*}! \, l_{k^*+1}!} \frac{\Gamma(e + \Delta_{N_{k^*}})}{(f + N_{k^*})^{e + \Delta_{N_{k^*}}}} \frac{f^e}{\Gamma(e)} \\
&\times \left(\prod_{i=1}^{N_{k^*}} \frac{1}{\Delta_{ik^*}!} \right) \frac{\Gamma(g+h)\Gamma(g + N_{k^*})\Gamma(h + N - N_{k^*})}{\Gamma(g)\Gamma(h)\Gamma(g+h+N)} \frac{d^c}{\Gamma(c)} \frac{\Gamma(c + l_{k^*})\Gamma(c + l_{k^*+1})}{\Gamma(c + l_k)(d+1)^{(c - \Delta_{k^*_{max}})}}.
\end{aligned} \tag{20}$$

The term T corresponds to the uniform sampling of τ_{2k^*-1} from the discrete time series interval, 2^N corresponds to sampling with equal probability a value of zero or

[10] Unless we currently have a new regime occurring in each period in which case $b_{T-1} = 0$.

one indicating the subset of series that are hit by the proposed break from a Bernoulli distribution and $(\Delta_{max}+1)^{N_{k^*}}$ corresponds to the uniform sampling of Δ_{ik^*} for $i \in N_{k^*}$ from the discrete interval $[0, \Delta_{max}]$.

The acceptance probability must be computed to ensure detailed balance is maintained across the entire parameter space including the number of breaks. We must therefore include in this acceptance probability, density terms that correspond to a hypothetical death move that would return the algorithm from the position proposed by the birth move to the current position. Therefore $(K+1)^{-1}$ corresponds to sampling with equal probability one of the $K+1$ breaks to remove. If the move is accepted we substitute all proposals for their original values.

3.4.2. Death move

With probability $d_k = 1 - b_k = 0.5$ a death move is entered which will attempt to remove the kth break window, which is sampled with equal probability from the existing set of K break windows, and therefore reduce K to $K^* = K - 1$.[11] This proposal entails removing τ_{2k-1} and τ_{2k}. We currently have two regimes k and $k+1$ separated by the break interval and are proposing to merge these into one longer regime denoted k^*. We compute $l_{k^*} = l_k + \Delta_{k_{max}} + l_{k+1}$ since the k^*th regime contains the observations $\tau_{2k-2}+1, \ldots, \tau_{2k+1}$. We further compute l^*, τ^*, ${\Sigma^*}^{-1}$, μ^*, $\tilde{a}^*$, $\tilde{b}^*$, and, for all i, l_i^* and τ_i^*. The move is accepted with probability $min(1, \alpha)$

$$\alpha = \frac{K\Gamma(e)\Gamma(a)}{T2^N(\Delta_{max}+1)^{N_k} f^e b^a} \left(\prod_{k=1}^{K^*+1} \frac{\mid \Sigma_k^* \mid^{1/2} \Gamma(\tilde{a}_k^*)}{{\tilde{b}_k^*}^{\tilde{a}_k^*}} \right) \left(\prod_{k=1}^{K+1} \frac{\tilde{b}_k^{\tilde{a}_k}}{\mid \Sigma_k \mid^{1/2} \Gamma(\tilde{a}_k)} \right) \mid V_\beta \mid^{1/2}$$
$$\times \frac{\Gamma(c)}{d^c} \frac{l_k!\, l_{k+1}!}{l_{k^*}!} \frac{\Gamma(c+l_{k^*})}{\Gamma(c+l_k)} \frac{(d+1)^{(c-\Delta_{k_{max}})}}{\Gamma(c+l_{k+1})} \frac{\Gamma(g)\Gamma(h)\Gamma(g+h+N)}{\Gamma(g+h)\Gamma(g+N_k)\Gamma(h+N-N_k)}$$
$$\times \frac{(f+N_k)^{e+\Delta_{N_k}}}{\Gamma(e+\Delta_{N_k})} \left(\prod_{i=1}^{N_k} \Delta_{ik}! \right). \tag{21}$$

The term K corresponds to sampling with equal probability one of the K break windows to remove. The term T^{-1} corresponds to the uniform sampling of τ_{2k-1} from the entire time series if we were hypothetically at $K-1$ breaks and proposing a birth move to return to our current position of K breaks. The 2^{-N} term corresponds to sampling with equal probability from a Bernoulli distribution a value of one or zero indicating which of the N series are hit by the break and the $(\Delta_{max}+1)^{-N_k}$ term

[11] Unless we currently have no breaks in which case $d_0 = 0$.

corresponds to sampling with equal probability from the break interval the lag with which each of the N_k series would be hit in the hypothetical proposed birth move. If the move is accepted we substitute all proposals for their original values and drop $\mathbb{1}_k$, N_k, Δ_{N_k}, δ_k, and p_k.

4. Hypothesis testing

This section presents the formal test of noncommon breaks. The framework developed in this article allows for noncommon breaks and is more complex than the nested common breaks model. It is therefore important to evaluate whether any improvement in fit from allowing for noncommon breaks is sufficient to warrant the increased model complexity. A Bayesian model evaluation is natural since it (i) integrates over all parameters and (ii) inherently penalises model complexity. Bayes Factors or marginal likelihoods are the preferred method for comparing models in a Bayesian framework (Kass and Raftery 1995). Bayes Factors, which determine the strength of evidence in favour of noncommon breaks, are constructed from the computed marginal likelihoods for the noncommon and common breaks models.

4.1. Hypothesis testing

We explain how to evaluate the strength of evidence in favour of (i) noncommon breaks and (ii) whether noncommonality is driven by nonpervasion or diffusion.

4.1.1. Hypothesis test for noncommon breaks

Chib (1995) shows that the marginal likelihood for model r can be computed as

$$p(\boldsymbol{y} \mid M_r) = \frac{p(\boldsymbol{y} \mid \theta^*, M_r)p(\theta^* \mid \boldsymbol{y})}{p(\theta^* \mid \boldsymbol{y}, M_r)} \tag{22}$$

in which θ^* denotes the posterior means of the parameter vector $\theta = (\boldsymbol{\beta}, \sigma^2)$. The calculation requires that the likelihood, prior and posterior are evaluated at this point.

Marginal likelihoods are computed for the models with noncommon breaks (M_{NCmn}) and common breaks (M_{Cmn}). Evidence is provided in favour of noncommon breaks through the Bayes Factor

$$BF_{Cmn} = \frac{p(\boldsymbol{y} \mid M_{NCmn})}{p(\boldsymbol{y} \mid M_{Cmn})}. \tag{23}$$

Kass and Raftery (1995) report that Bayes Factors between 1 and 3 are not worth a mention, between 3 and 20 indicate positive evidence in favour of noncommonality, between 20 and 150 indicate strong evidence, and even higher values very strong evidence.

4.1.2. Hypothesis test for nonpervasiveness of breaks

To evaluate whether any noncommonality is driven through the pervasion channel we compute the marginal likelihood for the model (M_{diff}) that restricts breaks to be fully pervasive ($p = 1$) but allows breaks to diffuse through the cross-section with different lags. Evidence in favour of the noncommon breaks model relative to this model will derive from the ability to allow a subset of series to be hit by breaks and is evaluated through the corresponding Bayes Factor

$$BF_{diff} = \frac{p(y \mid M_{NCmn})}{p(y \mid M_{diff})} \tag{24}$$

which will increasingly favour the noncommon breaks model the fewer series are hit. The greater the evidence in favour of the noncommon breaks model, the more of the noncommonality is driven through the pervasion channel.

4.1.3. Hypothesis test for diffuseness of breaks

We further compute the marginal likelihood for the model (M_{per}) which allows breaks to be less than fully pervasive but restricts all series that are hit to be so at the same date ($\delta = 0$ and $\Delta_{max} = 0$). The strength of evidence in favour of noncommon breaks relative to this model derives from the ability to allow breaks to be diffuse and is reflected in the corresponding Bayes Factor

$$BF_{per} = \frac{p(y \mid M_{NCmn})}{p(y \mid M_{per})}. \tag{25}$$

The strength of evidence in favour of noncommon breaks increases the more series are hit with delays. The more the noncommon breaks model outperforms this restricted model the more of the noncommonality is driven through the diffusion channel.

4.1.4. Which breaks drive noncommonality?

We can further evaluate (i) which breaks are driving any noncommonality and (ii) whether any noncommonality of the kth break is primarily driven by diffusion or a lack of pervasion. Since we have the posterior estimates of p_k and δ_k it is straightforward to evaluate this since $p_k = 1$ indicates the kth break is fully pervasive and $\delta_k = 0$ indicates it diffuses immediately through the entire cross-section.

5. Simulation study

To illustrate its ability to estimate noncommon breaks, the methodology developed here is compared with the common break models that have either pooled or unit-specific parameters developed by Smith and Timmermann (2017b) and the time series break model of Chib (1998) which is applied independently to each series in the cross-section.

5.1. Data generating process

The time series dimension T equals 200 and the cross-section dimension is gradually increased from $N = 10$ to $N = 50$. There are $K = 3$ breaks beginning at $\tau_1 = 50$, $\tau_2 = 100$ and $\tau_3 = 150$, respectively. Each break varies in its pervasiveness and diffuseness. The first break is assumed to be common and thus $p_1 = 1$ and $\delta_1 = 0$. The second hits every series with different lags such that $p_2 = 1$ and $\delta_2 = 3$. The third hits only a subset of series with different lags such that $p_3 = 0.6$ and $\delta_3 = 2$. The ith series is hit by the kth break if $\mathbb{1}_{ik}$=1 in which $\mathbb{1}_{ik} \sim$ Bernoulli(p_k). If $\mathbb{1}_{ik} = 1$ the ith series is hit by the kth break at time τ_{2k-1}+Δ_{ik} in which $\Delta_{ik} \sim Po(\delta_k)$. This allows us to construct $\boldsymbol{\tau}$.

For regimes $k = 1, \ldots, K + 1$, the dependent variable is constructed as

$$y_{it} = \beta_k' X_{it} + \epsilon_{it}, \qquad i \in N_{k-1}, \qquad t = \tau_{k_i - 1} + 1, \ldots, \tau_{k_i} \tag{26}$$

in which $X_{it} \sim N(0, 1)$. The slope β_k takes values of 1, 2.5, 2.5, and 4.75 in the four respective regimes. The errors ϵ_{it} are Normally distributed with zero mean and variance equal to 1 in the first two regimes and 2.5 in the final two.

5.2. Results

All models are estimated using hyperparameter values a=1, b=1, c=100, d=2, e=3, f=2, g=1, h=1, and $\sigma^2_\beta = 0.5$. These values imply that breaks occur on average every fifty periods, each series has equal prior probability of being hit by breaks and those series hit will be so with a prior expected length of three periods. The results are displayed in Table 1.

The time series break method detects just 4 of the 26 breaks that occur across all series and time periods. The left panel of Figure 2 displays, for the third break, the true break indicator values (blue circles) for each of the N series. The heterogeneous panel common break model (green crosses in the top left window) overlooks the third break, while the pooled common break model detects it but incorrectly assumes that every series is hit (red triangles).[12] Only the noncommon breaks model correctly identifies the six series that are hit (black crosses in bottom left window).

Table 2 displays the average across the ten series of the root mean squared error (RMSE) of each break estimate. The noncommon breaks model improves the third break estimate delivering a RMSE of 0.894 relative to the common break models with pooled (31.686) and unit-specific parameters (37.202).

The slope estimates from the various models are graphed in the right panel of Figure 2. The heterogeneous common break model (green dotted line in top right window) are biased in the third and fourth regimes for those series hit by the third break. It pools data across both regimes delivering a noisy estimate of $0.5 \times 2.5 + 0.5 \times 4.75 = 3.625$ which is biased 1.125 upwards in the third regime and 1.125 downwards in the fourth. The pooled common break model has biased slopes for all series in the final regime. It delivers a noisy estimate of $0.6 \times 4.75 + 0.4 \times 2.5 = 3.85$ which is biased 0.65 downwards for those series hit by the third break (red dashed line) and biased 1.35 upwards for the remaining series (red dashed line in bottom right window). The slope coefficients from the noncommon breaks model (black dashed lines) are unbiased for all series.

Table 1 displays that the regression coefficients are estimated accurately by the noncommon breaks model when $N = 10$ since their precision increases with the cross-section dimension and regime duration. Precision of the pervasion parameter p increases only with the cross-section. The left panel of Figure 3 displays the posterior distribution of the pervasion parameter p for the three breaks when $N = 10$. The

[12] These correspond to posterior modes of the break estimates.

vertical red lines mark the top and bottom 2.5% of the distribution. The estimates are close to the true values of 1, 1 and 0.6 thus correctly identifying that the third break is less pervasive than the first two. Increasing the cross-section to $N = 50$ results in more precise estimates of p (right panel).

The ability of the noncommon breaks model to capture the underlying data generating process is vastly improved compared with the common breaks models with pooled and unit-specific parameters. This is reflected by their respective Bayes Factors of 224.78 and 310.99. Since these values are above 150 they provide very strong evidence in favour of the noncommon breaks model relative to the common breaks models with pooled and heterogeneous coefficients.

6. Empirical Application

The empirical application explores the out-of-sample predictability of international excess stock returns using their corresponding dividend yields. Specifically, we compare the forecasting performance of our model that allows for noncommon breaks with a range of benchmark models that include two panel common break models with pooled and unit-specific parameters, respectively, a constant pooled panel model, a linear predictive regression applied independently to each country in the cross-section, and the country-level prevailing mean that acts as a 'no predictability' benchmark. Any improvement in predictive accuracy generated by our model is measured first in terms of statistical significance and second in terms of the economic utility a risk averse mean-variance investor would obtain from forecasting with our model relative to each of the benchmarks.

6.1. Data

The data set consists of $N = 17$ countries - Austria, Belgium, Canada, Denmark, Finland, France, Germany, Hong Kong, Italy, Japan, Malaysia, Netherlands, Singapore, Spain, Switzerland, U.K., and U.S. - in the cross-section and a worldwide series, the return on which acts as the first observed common factor. The second is the change in a global weighted exchange rate index. The latter is a trade-weighted average of a number of exchange rates against the U.S. dollar and thus proxies for the strength of the U.S. dollar.

Each monthly series begins in January 1973 and ends in November 2017 giving a time-series dimension $T = 539$. All data are from Global Financial Database.

Stock returns are computed from the closing index prices from 'Total Return Indices - Stocks'. Returns are constructed in excess of the corresponding country's three-month Treasury-bill rate. The dividend yield for each country in each month is computed using the average of dividends from month $t-11$ through month t.

As a robustness check we also generate forecasts using three popular alternative predictors: the three-month Treasury-bill rate, the term spread and the price-earnings ratio. When evaluating the predictive power of the price-earnings ratio, Austria, Finland, Italy, and Spain are omitted from our analysis due to data availability constraints. The term spread is measured as the difference between the yield on a country's ten-year government bond and the return on its three-month Treasury-bill rate. For this analysis we omit Hong Kong and Singapore due to lack of data. The price-earnings ratio and ten-year government bond data are also sourced from Global Financial Database.

We use value-weighted return indices throughout. Table 3 presents summary statistics for the monthly excess returns of each country. The average monthly excess returns range from 0.08 for Germany to 1.05 for Hong Kong. The minimum and maximum values combined with the high standard deviations reflect the high volatility in the series, with Hong Kong, Malaysia and Singapore being the most volatile. Finland, Hong Kong, Netherlands, Switzerland, and United States all have Sharpe Ratios above 0.10 with Germany having the lowest value of 0.01. A handful of countries have relatively large positive autocorrelations such as Austria (0.21) and Finland (0.25), while United States has the smallest autocorrelation (0.03).

6.2. Model Specifications

Our noncommon breaks model, for $i \in N_{k-1}$, $t = \tau_{k-1}+1, \ldots, \tau_k$ and $k = 1, \ldots, K+1$, is specified as

$$y_{it} = \beta_{0k} + \beta_{1k} X_{it-1} + \gamma_{1k} f_{1t} + \gamma_{2k} f_{2t} + e_{it} \tag{27}$$

in which y_{it} is the return on the ith country's aggregate stock market at time t in excess of its risk-free rate and X_{it-1} denotes the log dividend yield for the ith country at time t. The two observed common factors at time t, the return on a global stock market portfolio and the change in a global exchange rate index, are denoted by f_{1t} and f_{2t}. Their loadings are denoted γ_{1k} and γ_{2k}.

The first benchmark model is a constant pooled panel model. For $i = 1, \ldots, N$

and $t = 1, \ldots, T$, the model is specified as

$$y_{it} = \beta_0 + \beta_1 X_{it-1} + \gamma_1 f_{1t} + \gamma_2 f_{2t} + e_{it}. \tag{28}$$

The second benchmark is a pooled panel model with common breaks. For $i = 1, \ldots, N$, $t = \tau_{k-1} + 1, \ldots, \tau_k$ and $k = 1, \ldots, K+1$, this model is specified as

$$y_{it} = \beta_{0k} + \beta_{1k} X_{it-1} + \gamma_{1k} f_{1t} + \gamma_{2k} f_{2t} + e_{it}. \tag{29}$$

The third is a panel common breaks model with unit-specific parameters. For $i = 1, \ldots, N$, $t = \tau_{k-1} + 1, \ldots, \tau_k$ and $k = 1, \ldots, K+1$, the model is specified as

$$y_{it} = \beta_{i0k} + \beta_{i1k} X_{it-1} + \gamma_{i1k} f_{1t} + \gamma_{i2k} f_{2t} + e_{it}. \tag{30}$$

A time series linear regression estimated using Ordinary Least Squares applied independently to each country in the cross-section is the fourth benchmark model. The model specification for $i = 1, \ldots, N$ and $t = 1, \ldots, T$ is

$$y_{it} = \beta_{0i} + \beta_{1i} X_{it-1} + e_{it}. \tag{31}$$

Finally, forecasts are generated using the prevailing mean in each country which serves as a 'no predictability' benchmark.

6.3. Testing cross-sectional dependence

International stock returns are likely to be highly correlated across the cross-section. To formally evaluate the degree of correlation we compute the average pairwise correlation across the series and conduct the cross-sectional dependence (CD) test of Pesaran (2004) which is well suited because it is robust to multiple breaks. Using the full sample of data we first estimate pairwise correlations $\hat{\rho}_{ij}$ of the residuals from series i and j in which $i \neq j$

$$\hat{\rho}_{ij} = \frac{\sum_{t=1}^{T} e_{it} e_{jt}}{\left(\sum_{t=1}^{T} e_{it}^2\right)^{1/2} \left(\sum_{t=1}^{T} e_{jt}^2\right)^{1/2}} \tag{32}$$

where e_{it} is the residual that is obtained from an Ordinary Least Squares time series regression for series i

$$e_{it} = y_{it} - \hat{\beta}_{0i} - \hat{\beta}_{1i} X_{it-1}, \qquad t = 2, \ldots, T. \tag{33}$$

The pairwise correlations are used to compute the CD test statistic which has a standard Normal distribution

$$CD = \sqrt{\frac{2T}{N(N-1)}} \left(\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \hat{\rho}_{ij} \right). \tag{34}$$

The computed CD statistic is equal to 126.028 conclusively rejecting the null hypothesis of weak cross-sectional dependence. The mean of the absolute pairwise correlations is 0.465 reflecting strong cross-sectional dependence in the data.

Including the two observed common factors in the regression in equation (33) the computed CD statistic is 24.510 and the average absolute pairwise correlation is 0.086 suggesting that the common factors have removed approximately 80% of the cross-sectional dependence. The ability of the method to detect breaks will likely be unaffected since Smith et al. (2017) report that break detection power only diminishes once the cross-sectional error correlations rise above 0.2.

Finally, our empirical application does not attempt to perform inference on the parameter estimates. The ultimate test of our model is whether it can generate more accurate forecasts than a range of benchmark models that include the consistently estimated linear model.

6.4. Priors

In the following empirical application it is assumed a priori that a break occurs every ten years for the break models. This is specified by setting the hyperparameters c=240 and d=2. The error-term variance is specified through the hyperparameters a and b which are set equal to 2. For the noncommon break model it is assumed that each series has equal prior probability of being hit by any given break. This belief is achieved by setting $g = h = 1$. Setting the hyperparameters $f = 1$ and $e = 3$ implies that a priori each series is hit by any given break with a delay of three periods.

The key priors are motivated by economics. The prior on the slope coefficient β_k is a Gaussian distribution centered on zero because such a model corresponds to no predictability as documented by Goyal and Welch (2008). The variance σ_β^2 of the distribution controls the degree of predictability that investors believe resides in the market. If $\sigma_\beta^2 = 0$ investors have a dogmatic belief that returns are not predictable and thus the R^2 from the predictive regression equals zero, whereas if $\sigma_\beta^2 = \infty$ investors believe any degree of predictability is equally likely. Positive but finite values of σ_β^2 imply investors are sceptical about the degree of predictability but do not altogether rule it out.

Wachter and Warusawitharana (2009) note that if the predictive variable has a large variance σ_X^2 the prior on β will likely be lower while if the error-term variance σ^2 is high the prior on β will likely be higher. In our setting, for the kth regime it is desirable to scale β_k by $\sigma_{X,k}^2$ and σ_k^2 and place the prior over this 'normalised beta'

$$\eta_k = \beta_k \frac{\sigma_{X,k}}{\sigma_k}, \qquad k = 1, \dots, K+1. \tag{35}$$

The prior on η_k is

$$p(\eta_k) \sim N(0, \sigma_\eta^2), \qquad k = 1, \dots, K+1. \tag{36}$$

This is equivalent to placing the following prior on β_k

$$p(\beta_k) \sim N\left(0, \frac{\sigma_\eta^2}{\sigma_{X,k}^2} \sigma_k^2\right), \qquad k = 1, \dots, K+1 \tag{37}$$

in which $\sigma_{X,k}^2$ is the empirical variance of the predictor computed across the N series in the kth regime.

Placing the prior on this 'normalised beta' is appealing since it maps directly into a prior on the R^2 from the predictive regression and thus the prior degree of scepticism surrounding the amount of predictability in the market is defined by σ_η. A value of $\sigma_\eta = 0$ implies investors are dogmatic in their belief that returns are unpredictable. As the value of $\sigma_\eta = 0$ increases investors become less sceptical about the existence of return predictability and thus allocate more probability to positive R^2 values. Figure 4 displays that when $\sigma_\eta = 0.04$ investors assign 0.075 probability to R^2 values greater than 0.005, whereas large values of σ_η imply that investors assign equal probabilities to all possible R^2 values.

The empirical application assumes a moderate degree of predictability defined by setting $\sigma_\eta = 0.04$ as suggested by Wachter and Warusawitharana (2009). We also evaluate the robustness of the results to other values of σ_η.

6.5. Break dynamics

Our discussion focuses on the model that generates forecasts with the dividend yield as the predictive variable alongside an intercept.

The top window of Figure 5 displays that the noncommon breaks model places 86% of the posterior probability on four breaks occurring within the 45 year sample period corresponding to a break occurring approximately every ten years. The break dates (bottom window) correspond to key economic events such as the oil price shocks of the 1970s, the stock market crash in the early 2000s and the global financial crisis.

When the predictive variable is the dividend yield (dy), the financial crisis is the most pervasive break with a posterior mean of 0.958 for its pervasion parameter p_4 (left column of Table 4), implying that the crisis was transmitted across the world. The posterior modes of the break indicators displayed in Figure 6 (bottom right window) show that every country was hit by the financial crisis. Both oil price shocks (top windows) were also very pervasive hitting all but Hong Kong, Malaysia and Singapore resulting in posterior means of p_1 and p_2 equal to 0.796 and 0.828, respectively. Ferson and Harvey (1994) report that the equity markets of these three countries, unlike the majority of countries, are positively related to oil price changes which might explain why they are not hit by these breaks. The least pervasive break was in 2003 which hit the Group of Seven (G7) countries and Hong Kong, Malaysia and Singapore.

Figure 6 displays that the U.S. is the first country to be hit by all four breaks with other countries being hit with a 0-4 month lag. This supports the notion that macroeconomic shocks either originate in, or first hit, the U.S. and are subsequently transmitted to other countries fairly quickly. With a news diffusion model, Rapach et al. (2013) report that shocks to the U.S. propagate to other countries with a delay which gives rise to predictability of international stock returns using lagged U.S. returns.[13]

The posterior means of the diffusion parameter δ for all four breaks displayed in Table 4 imply that all shocks diffused rapidly through the economy with the second oil price shock taking slightly longer than the others. The estimates of p and δ suggest that there are some international frictions with shocks not transmitting to all parts of the economy, but when transmission does take place it does so rapidly.

We now formally test whether noncommon breaks are preferred relative to common breaks. Bayes Factors of 141.77 and 282.29 provide strong and very strong evidence, respectively, in favour of noncommon breaks relative to panel common break models with pooled and unit-specific parameters. Of the two common breaks models, the one with pooled parameters fares better.

[13] See Hong et al. (2007) for an application of the news diffusion model to U.S. industries.

6.6. Group structure

A group structure that induces cross-sectional heterogeneity is inferred from the pervasiveness of breaks. Table 5 displays the groups. Hong Kong, Malaysia and Singapore form their own group since they are the only countries not hit by the oil price shocks. Furthermore, the G7 countries form their own group as they are hit by all four breaks while the remaining countries - Austria, Belgium, Denmark, Finland, Netherlands, Spain, and Switzerland - are not hit by the break in the early 2000s.

Table 3 displays that Hong Kong, Malaysia and Singapore exhibit the highest volatility of all the countries in the sample. This group is therefore characterised by high volatility, while the second and third groups that have lower volatility are separated by their level of income and development: the G7 countries are more developed than the remaining seven countries.

6.7. Out-of-sample return forecasts

Using an initial estimation period of ten years, forecasts are recursively generated at each point in the out-of-sample period using only the information available at the time the forecast is made. We first focus on the evolution in the U.S. return forecasts displayed in Figure 7.

The historical average (solid black line in bottom window) provides the smoothest forecasts of all. The forecasts from the pooled panel common break model (solid black line in middle window) span the narrowest range which results from parameters being pooled across the entire cross-section. The remaining models are all able to generate forecasts that can span a broader range of both negative and positive forecasts, reflecting the importance of allowing for cross-sectional heterogeneity. By the same token, the heterogeneous panel common break model (red dashed line in middle window) and the linear time series model (red dashed line in bottom window) display more volatility during the early part of the out-of-sample period compared with the noncommon breaks model (top window). This indicates the benefit of exploiting information in the cross-section by pooling parameters particularly at the beginning of the out-of-sample period when the available time-series dimension is relatively short.

6.7.1. *Evolution of return forecasts*

The comparative forecasting performance of our model relative to each of the benchmarks is measured for the ith country by the popular out-of-sample R_i^2 value

$$R_i^2 = 1 - MSE_{i,NCmn}/MSE_{i,Bmk} \tag{38}$$

in which $MSE_{i,NCmn}$ and $MSE_{i,Bmk}$ denote the mean squared forecast errors for the ith country obtained from our noncommon breaks model and the given benchmark model. Positive R_i^2 values indicate our method is outperforming the benchmark, while negative values reflect underperformance. Campbell and Thompson (2008) note that for monthly data an R^2 value as small as 0.5% reflects strong outperformance.

Figure 8 displays the distribution of R_i^2 values across the 17 countries. Each window corresponds to one of the five benchmark models. Relative to the time series linear model (top left window), our model outperforms for 16 out of 17 countries with many of the values above 0.5%. For the U.S. (marked by the thick vertical black line) the R^2 value is approximately 1%.

Equally strong outperformance is measured relative to the constant pooled panel model (top right), the panel common breaks models with pooled and unit-specific parameters (middle windows) and the historical average (bottom left). Across all benchmarks, our model only underperforms for one country, frequently delivers R^2 values greater than 0.5% and always outperforms for the U.S. delivering R^2 values between 0.5% and 1%.

6.7.2. *Statistical significance of improved predictive accuracy*

To evaluate whether any out- or under-performance of our model in forecast accuracy is statistically significant, we first implement the mean squared error differential tests proposed by Diebold and Mariano (1995). When comparing the predictive accuracy of nested models, however, the $t-$statistic may have a nonstandard distribution. To account for this possible nonstandard distribution, we also implement the MSE-adjusted test statistic proposed by Clark and West (2007).

The results are displayed in Table 6. The table displays the number of countries for which our model significantly underperforms ($t < -1.64$), insignificantly underperforms ($-1.64 < t < 0$), insignificantly outperforms ($0 < t < 1.64$), and significantly outperforms ($t > 1.64$) the benchmark according to both the Diebold-Mariano (DM) and Clark-West (CW) tests. The five panels correspond to each of the benchmark models. The top row of each panel presents results when predicting with the dividend yield (dy). The † symbol denotes the bin in which the U.S. falls.

Relative to the time series linear model, our model significantly outperforms for 15 out of 17 countries including the U.S.. Our model also outperforms for another country but the outperformance is insignificant and it never significantly underperforms. The statistical significance of outperformance remains strong across the other four benchmarks with our model significantly outperforming for between 13 and 16 countries, always significantly outperforming for the U.S. and never significantly underperforming.

Turning to the Clark-West statistic, the results are marginally stronger. We significantly outperform for between 14 and 16 countries, always significantly outperform for the U.S. and never significantly underperform.

6.7.3. Consistent improvements in predictive accuracy

Up to this point, the evaluation of outperformance has been averaged across the out-of-sample period. However, a strong forecasting model should perform well across the majority of the out-of-sample period and not only around one or two short periods. To evaluate whether our model out- or under-performs each of the benchmarks consistently across the out-of-sample period we compute the cumulative sum of squared error differences ($CSSED$) measure for countries $i = 1, \ldots, N$

$$CSSED_{it} = \sum_{\tau=1}^{\tau=t}(e^2_{Bmk,i\tau} - e^2_{NCmn,i\tau}), \qquad t = m+1, \ldots, T \tag{39}$$

in which $m = 120$ denotes the initial estimation period of ten years and $e_{Bmk,i\tau}$ and $e_{Ncmn,i\tau}$ denote the forecast errors from the benchmark model and the noncommon breaks model for the ith country at time τ.

If our model is outperforming we will see positive and rising values of the $CSSED$ measure while if our model is underperforming we will see negative and declining values. Any improved performance that is focused around one or two short periods will be reflected in large upward spikes around those periods with flat spots, which reflect equal predictive performance of the two models, elsewhere. A gradual but irregular upward-sloping curve suggests our model outperforms consistently across the sample.

Figure 9 graphs the $CSSED$ values relative to the five benchmark models. The four windows correspond to U.S., Italy, U.K., and Germany. For the U.S. (top left window), our model outperforms the historical average (black solid line) consistently through the sample with only a few short periods of underperformance. This strong and consistent outperformance is striking given how difficult the historical average has proven to beat out-of-sample (Goyal and Welch 2008). The outperformance is similarly strong and consistent against the other benchmark models.

The outperformance continues to be consistent across the out-of-sample period for Germany, Italy and U.K.. The only notable period of prolonged underperformance is relative to the pooled panel common break model (blue dashed line) during the mid-2000s.

6.8. Economic utility from return forecasts

Having evaluated whether the more accurate forecasts generated by the noncommon breaks model relative to the benchmark models are statistically significant, it is important to evaluate whether they are economically meaningful. We conduct two utility exercises. First, for each of the 17 countries a risk averse mean-variance investor allocates his wealth in each period between the risk-free rate and the risky asset of that country (Goyal and Welch 2008; Campbell and Thompson 2008; Rapach et al. 2010). Using the noncommon breaks model, at time t the investor allocates a weight to equities of the ith country in period $t+1$ equal to

$$w_{NCmn,i,t} = \frac{1}{A}\frac{\hat{r}_{NCmn,i,t+1}}{\hat{\sigma}^2_{Stk,i,t+1}}, \tag{40}$$

in which $\hat{\sigma}^2_{Stk,i,t+1}$ is the stock variance estimated for the ith country at period $t+1$ using only information available at time t. Following Campbell and Thompson (2008), we use a rolling window of 60 monthly stock returns to estimate this variance, assume a risk aversion coefficient of $A = 3$ and truncate the weight in the risky asset at 0% and 150% to prohibit short-selling and excessive leverage. Investing with the noncommon breaks model the utility gain to the investor for the ith country is

$$\hat{\upsilon}_{NCmn,i} = \hat{\mu}_{NCmn,i} - \frac{A\hat{\sigma}^2_{NCmn,i}}{2}, \tag{41}$$

in which $\hat{\mu}_{NCmn,i}$ and $\hat{\sigma}^2_{NCmn,i}$ denote the sample mean and variance of the portfolio constructed from the forecasts of the noncommon breaks model throughout the out-of-sample period.

We compute the average utility for each of the benchmark models in the same way whereby the investor allocates the following weight to equities of the ith country at time t

$$w_{Bmk,i,t} = \frac{1}{A}\frac{\hat{r}_{Bmk,i,t+1}}{\hat{\sigma}^2_{Stk,i,t+1}} \tag{42}$$

capturing an average utility of

$$\hat{v}_{Bmk,i} = \hat{\mu}_{Bmk,i} - \frac{A\hat{\sigma}^2_{Bmk,i}}{2}, \tag{43}$$

in which $\hat{\mu}_{Bmk,i}$ and $\hat{\sigma}^2_{Bmk,i}$ denote the sample mean and variance of the portfolio constructed using the forecasts from the benchmark model.

Multiplying the difference between equations (41) and (43) by 1200 obtains the annualised percentage certainty equivalent return (CER) for the ith country obtained from forecasting with the noncommon breaks model relative to the benchmark. This calculation is performed for each of the 17 countries.

Figure 10 displays the distribution of the CER values across the 17 countries obtained from the forecasts generated from the noncommon breaks model relative to the benchmark when using the dividend yield. Each of the five windows correspond to a different benchmark.

The plots show that our model delivers a positive CER for at least 14 out of 17 countries across all five benchmarks. The magnitudes of the CER values are economically large for almost all countries with the U.S. having a CER value approximately between 1.5% and 2% against each benchmark.

6.8.1. Industry allocation analysis

The second exercise evaluates the utility gain to a risk averse mean-variance investor who recursively allocates his wealth between the (U.S.) risk free rate and a risky portfolio which is constructed from the risky assets of the 17 countries (see e.g. Avramov and Wermers (2006), Banegas et al. (2013) and Smith and Timmermann (2017a)).

The return on the risky portfolio in excess of the U.S. risk-free rate at time $t+1$ is denoted $\tilde{r}_{p,t+1}$. This excess return is equal to the sum of the aggregate stock returns of the 17 countries multiplied by a value weight vector w_{t+1} minus the U.S. risk-free rate, $r_{f,t,US}$. The weight vector is determined using our noncommon breaks model at each time t. Specifically, the weights are determined using numerical optimisation techniques that maximise the expected utility function

$$E[U(\tilde{r}_{p,t+1} \mid A)] = r_{f,t,US} + w_t'\hat{r}_{t+1} - \frac{A}{2}w_t'\hat{S}_t w_t, \tag{44}$$

subject to the constraints that (i) the portfolio weights sum to one $\sum_{i=1}^{N} w_{it} = 1$ and (ii) there is no short selling or leverage of any country's risky asset $w_{it} \in [0, 1]$ for $i = 1, \dots, N$.

Constraining portfolio weights is similar in spirit to shrinkage techniques applied directly to the variance-covariance matrix which has proven successful in improving performance in mean-variance analysis (Jagannathan and Ma 2003; DeMiguel et al. 2007). For each time point at which forecasts are recursively generated, the covariance matrix, $\hat{S}_t$, is estimated using the residuals from the predictive regression using only the data available at that time.

The realised utility at time t is computed by plugging w_t back into equation (44). This computation is conducted recursively throughout the out-of-sample period for all the return prediction models we consider. The CER is expressed as an annualised percentage since we multiply by 1200 the difference between the utility from the noncommon breaks model in equation (44) and the same utility calculated from the forecasts of the benchmark model.

Table 7 displays for every model the average allocation through time to each of the 17 countries. The average weight allocations vary across the six models from which forecasts are generated. For instance, the average allocation to Hong Kong is as low as 2% under the linear time series model but nearly 20% under the noncommon breaks model. The historical average allocates the largest amount to Germany (23%) while the linear time series model allocates the largest average weight to the U.K. (21%). The noncommon breaks model also allocates considerable average weights to Malaysia (17%) and Singapore (16%).

Table 8 displays the utility gains to a mean-variance investor with moderate risk aversion who at each period allocates his wealth between the U.S. risk-free rate and a risky portfolio constructed from the optimized weight allocations across the risky assets of the 17 countries. Specifically, the table reports the CER values expressed as annualised percentages from forecasting with the noncommon breaks model relative to the five benchmarks when the dividend yield (dy) is the predictive variable (top row). Relative to the pooled panel model with common breaks, the CER value is 3.23% per year. Against the remaining four benchmarks, the CER value remains above 2% which is economically significant.

The large CER values imply that a mean-variance investor with moderate risk aversion could have derived economically meaningful utility gains from generating forecasts in real time from the noncommon breaks model relative to each of the five benchmarks.

7. Extensions and robustness

This section evaluates the robustness of the results. First, results are presented for three other popular predictors of excess stock returns: the T-bill rate, the term spread and the price-earnings ratio. Second, we investigate the sensitivity of the baseline results to the choice of priors by adjusting one prior hyperparameter at a time. Finally, we suggest some possible extensions of the methodology.

7.1. Results for alternative predictors

The results presented thus far were produced from a predictive regression that included an intercept term and lagged country-level dividend yields as the predictive variable. We now present results when forecasting with other predictors that have been proposed by the literature on stock return predictability. Specifically, we present results when the predictor is either the T-bill rate, term spread or price-earnings ratio.[14] Consistent outperformance across these other predictors will instill confidence in the broad applicability of the method.

We generate forecasts recursively in the same manner as before using an initial estimation period of ten years. To save space we present less detailed results than for the dividend yield. First, the noncommon breaks model detects four breaks in the sample period when using each of the three alternative predictive variables. Across these three additional predictors we find extensive evidence of breaks being nonpervasive and diffuse (Table 4).

The improvement in predictive accuracy delivered by the noncommon breaks model continues to hold when forecasting with the T-bill rate (tbl), term spread (tms) or the price-earnings ratio (pe). Table 6 displays the number of countries for which our method significantly underperforms, insignificantly underperforms, insignificantly outperforms, or significantly outperforms. Across all five benchmarks, the noncommon breaks model generates significantly more accurate forecasts for the majority of countries using either the Diebold-Mariano or Clark-West test. Our model only significantly underperforms once when the price-earnings ratio is the predictive variable.

[14] Recall from Section 6.1 that due to data availability constraints we only have 13 and 15 countries when forecasting with the price-earnings ratio and term spread.

The significantly more accurate forecasts also translate into economic utility gains for a mean-variance investor with moderate risk aversion. Table 8 displays the CER values relative to each of the five benchmarks when forecasting with the alternative predictors. Forecasting with the noncommon breaks model continues to generate annualised utility gains in the region of 2-3% relative to each benchmark and across all predictors.

The results for these three additional predictors are in line with the more detailed analysis for the dividend yield suggesting our findings are not driven by the choice of predictor.

7.2. Sensitivity of results to choice of priors

We now explore the sensitivity of the baseline results generated with the dividend yield as the predictive variable to the choice of prior by adjusting one hyperparameter at a time. Table 9 displays the results when we substitute the hyperparameter value reported in the table for its original value while allowing all remaining hyperparameter values to be unchanged from their values detailed in Section 6.4.

Changing our prior expected regime duration from 20 to 10 years ($c = 120$) leads to five breaks being detected. This prior adjustment does not affect the ability of the noncommon breaks model to generate significantly more accurate forecasts than the five benchmarks for the majority of the 17 countries. Setting the prior expected delay with which series are hit by breaks to five periods ($e = 5$) and increasing the prior probability of series being hit by each break to 90% ($g = 10$) does not affect the forecasting performance of the noncommon breaks model. Finally, the outperformance is robust to changes in the prior volatility (controlled by b) and the degree of scepticism regarding how much predictability exists in the market (controlled by σ_η^2).

7.3. Partial break models

The current framework assumes that all parameters, the intercept, slope coefficients and error-term variance, follow the same breakpoint process. In practice, such an assumption may be too restrictive. Peluso et al. (2016) (see also Bai and Perron (1998)) explain how to specify multiple breakpoint processes that separately affect different regression parameters. For instance, we may specify one breakpoint process that affects the slope coefficient and thus controls how return predictability varies through time and another that controls how the volatility varies through time. This is particularly appealing if there are economic reasons to suggest that different break-point processes control return predictability and volatility. For example, one may specify a break process driven by financial crises that drives shifts in volatility while it may be changes in monetary policy that drive shifts in return predictability. A full exposition of such a model is beyond the scope of this paper.

7.4. Fixed effects

We can further include in the model additive individual fixed effects that do not vary through time. Bonhomme and Manresa (2015) report that such a model requires time-variation in the covariates X_{it}. If lagged dependent variables are incorporated into the panel regression the inclusion of individual fixed effects will give rise to the well documented Nickell (1981) bias. If one is only interested in forecasting this may be unproblematic but if inference on the parameter estimates is being performed one would need to implement an estimation procedure, such as instrumental variables, to obtain consistent parameter estimates.

8. Conclusion

This article develops a new Bayesian approach to estimate noncommon structural breaks in panel regression models. Breaks can hit any subset of series in the cross-section at different times within a break window. Parameters are pooled across the cross-section to reduce parameter estimation error but nonpervasive breaks lead to series being in different regimes at a given time thus introducing cross-sectional heterogeneity. We develop a formal test of noncommon breaks and whether any noncommonality is driven primarily through the pervasion or diffusion channel. Which breaks drive noncommonality is revealed by the pervasion and diffusion parameters that are learned from the cross-section.

In an empirical application that recursively generates out-of-sample forecasts of monthly excess stock returns for 17 countries using their corresponding lagged country-level dividend yields for the sample period 1983 through 2017, we detect four breaks at 1975, 1979, 2002, and 2008. There is evidence of noncommonality through both nonpervasion and diffusion across all four of these breaks. The U.S. is hit first by each break supporting the evidence of Rapach et al. (2013) provided through a news diffusion model that large macroeconomic shocks either originate in, or first hit, the U.S. and subsequently propagate to other countries causing U.S. returns to be a stronger predictor of the returns of other countries than their own lagged dividend yields. The noncommon breaks model generates more accurate forecasts relative to a range of popular benchmark models including, among others, the panel common break models with either pooled or unit-specific parameters and the historical average. The improved predictive accuracy is consistent across the out-of-sample period and is both statistically significant and economically meaningful.

The practical value of the method derives from its many economic applications. The methodology could be readily applied to estimation of news diffusion models that give rise to lead-lag predictability (Hong et al. 2007), analysing the pace of diffusion of technological innovations through the economy (Kapur 1995), the propagation of macroeconomic shocks or recessions across countries or regions (Hamilton and Owyang 2012), and revealing more closely the true nature of 'pockets' of return predictability (Farmer et al. 2018). The diffusion of shocks becomes more relevant the higher the frequency of data being studied and this is of particular interest given the recent surge in the analysis of high frequency data particularly in financial markets.

Appendix A: marginalizing $\boldsymbol{\beta}$ and σ^2 from the posterior

This Appendix shows how we marginalize the regression parameters $\boldsymbol{\beta}$ and σ^2 from the posterior and therefore how (14) is derived from (13). If we simply rewrite (13) and multiply and divide it by $2\pi^{-\kappa/2} \mid \Sigma_k \mid^{-1/2} (\sigma_k^2)^{-\kappa/2}$ we can first marginalize β using (15)

$$p(\sigma^2 \mid \boldsymbol{y}, \boldsymbol{X}, \boldsymbol{\tau}) = \int p(\sigma^2, \boldsymbol{\beta} \mid \boldsymbol{y}, \boldsymbol{X}, \boldsymbol{\tau}) d\boldsymbol{\beta}$$
$$= \left(\prod_{k=1}^{K+1} 2\pi^{-(\not{\kappa}+l_{N_{k-1}})/2} (\sigma_k^2)^{-(a+1+\not{\kappa/2}+l_{N_{k-1}}/2)} \frac{b^a}{\Gamma(a)} \mid V_\beta \mid^{-1/2} \mid \Sigma_k \mid^{1/2} \cancel{2\pi^{\kappa/2} (\sigma_k^2)^{\kappa/2}} \right)$$
$$\times\, exp \left[- \sum_{k=1}^{K+1} \frac{1}{2\sigma_k^2} \left(2b + \mu_k' \Sigma_k^{-1} \mu_k + \sum_{i=1}^{N_{k-1}} \sum_{t=\tau_{k_i-1}+1}^{\tau_{k_i}} y_{it}^2 \right) \right]$$
$$\cancel{\times \int \prod_{k=1}^{K+1} 2\pi^{-\kappa/2} (\sigma_k^2)^{-\kappa/2} \mid \Sigma_k \mid^{-1/2} exp \left[\sum_{k=1}^{K+1} \frac{-1}{2\sigma_k^2} (\beta_k - \mu_k)' \, \Sigma_k^{-1} \, (\beta_k - \mu_k) \right] d\beta_k} \tag{A.1}$$

whereby the final term cancels since the full conditional distribution of β_k is Gaussian with mean vector μ_k and covariance matrix $\Sigma_k \sigma_k^2$ and hence its integral equals one.

We further marginalize σ^2 by rearranging terms in (A.1) and multiplying and dividing by $\Gamma(\tilde{a}_k)/\tilde{b}_k^{\tilde{a}_k}$

$$p(\boldsymbol{y} \mid \boldsymbol{X}, \boldsymbol{\tau}) = \int p(\sigma^2 \mid \boldsymbol{y}, \boldsymbol{X}, \boldsymbol{\tau}) d\sigma^2 = \left(\prod_{k=1}^{K+1} 2\pi^{-l_{N_{k-1}}/2} \frac{b^a}{\Gamma(a)} \mid V_\beta \mid^{-1/2} \mid \Sigma_k \mid^{1/2} \frac{\Gamma(\tilde{a}_k)}{\tilde{b}_k^{\tilde{a}_k}} \right)$$
$$\cancel{\times \int \left(\prod_{k=1}^{K+1} \frac{\tilde{b}_k^{\tilde{a}_k}}{\Gamma(\tilde{a}_k)} (\sigma_k^2)^{-(\tilde{a}_k+1)} \right) exp \left[- \sum_{k=1}^{K+1} \frac{\tilde{b}_k}{\sigma_k^2} \right] d\sigma_k^2} \tag{A.2}$$

whereby the canceling arises because the full conditional distribution of σ_k^2 is an inverse gamma with parameters $\tilde{a}_k$ and $\tilde{b}_k$ and thus its integral equals one.

References

Andrews, D. W., "Tests for parameter instability and structural change with unknown change point," *Econometrica* 61 (1993), 821–856.

———, "End-of-sample instability tests," *Econometrica* 71 (2003), 1661–1694.

———, "Cross-section Regression with Common Shocks," *Econometrica* 73 (2005), 1551–1585.

Andrews, D. W. and W. Ploberger, "Optimal tests when a nuisance parameter is present only under the alternative," *Econometrica* 62 (1994), 1383–1414.

Avramov, D. and R. Wermers, "Investing in mutual funds when returns are predictable," *Journal of Financial Economics* 81 (2006), 339–377.

Bai, J., "Panel data models with interactive fixed effects," *Econometrica* 77 (2009), 1229–1279.

———, "Common breaks in means and variances for panel data," *Journal of Econometrics* 157 (2010), 78–92.

Bai, J. and J. L. Carrion-I-Silvestre, "Structural changes, common stochastic trends, and unit roots in panel data," *Review of Economic Studies* 76 (2009), 471–501.

Bai, J., R. L. Lumsdaine and J. H. Stock, "Testing for and dating common breaks in multivariate time series," *Review of Economic Studies* 65 (1998), 395–432.

Bai, J. and P. Perron, "Estimating and testing linear models with multiple structural changes," *Econometrica* 66 (1998), 47–78.

———, "Computation and analysis of multiple structural change models," *Journal of Applied Econometrics* 18 (2003), 1–22.

Baltagi, B. H., Q. Feng and C. Kao, "Estimation of heterogeneous panels with structural breaks," *Journal of Econometrics* 191 (2016), 176–195.

Baltagi, B. H., C. Kao and L. Liu, "Estimation and identification of change points in panel models with nonstationary or stationary regressors and error term," *Econometric Reviews* 36 (2017), 85–102.

Banegas, A., B. Gillen, A. Timmermann and R. Wermers, "The cross section of conditional mutual fund performance in European stock markets," *Journal of Financial Economics* 108 (2013), 699–726.

Bonhomme, S. and E. Manresa, "Grouped patterns of heterogeneity in panel data," *Econometrica* 83 (2015), 1147–1184.

Campbell, J. Y. and S. B. Thompson, "Predicting excess stock returns out of sample: Can anything beat the historical average?," *Review of Financial Studies* 21 (2008), 1509–1531.

Carlin, B. P. and S. Chib, "Bayesian model choice via Markov chain Monte Carlo methods," *Journal of the Royal Statistical Society. Series B (Methodological)* 57 (1995), 473–484.

Chen, B. and Y. Hong, "Testing for smooth structural changes in time series models via nonparametric regression," *Econometrica* 80 (2012), 1157–1183.

Chib, S., "Marginal likelihood from the Gibbs output," *Journal of the American Statistical Association* 90 (1995), 1313–1321.

———, "Estimation and comparison of multiple change-point models," *Journal of Econometrics* 86 (1998), 221–241.

Chib, S. and K. H. Kang, "Change-points in affine arbitrage-free term structure models," *Journal of Financial Econometrics* 11 (2012), 302–334.

Clark, T. E. and K. D. West, "Approximately normal tests for equal predictive accuracy in nested models," *Journal of Econometrics* 138 (2007), 291–311.

De Wachter, S. and E. Tzavalis, "Monte Carlo comparison of model and moment selection and classical inference approaches to break detection in panel data models," *Economics Letters* 88 (2005), 91–96.

———, "Detection of structural breaks in linear dynamic panel data models," *Computational Statistics & Data Analysis* 56 (2012), 3020–3034.

DeMiguel, V., L. Garlappi and R. Uppal, "Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy?," *Review of Financial studies* 22 (2007), 1915–1953.

Diebold, F. X. and R. S. Mariano, “Comparing predictive accuracy,” *Journal of Business & Economic Statistics* 13 (1995), 253–263.

Elliott, G. and U. K. Müller, “Efficient tests for general persistent time variation in regression coefficients,” *Review of Economic Studies* 73 (2006), 907–940.

Farmer, L., L. Schmidt and A. Timmermann, “Pockets of predictability,” *University of California San Diego Manuscript* (2018).

Ferson, W. E. and C. R. Harvey, “Sources of risk and expected returns in global equity markets,” *Journal of Banking & Finance* 18 (1994), 775–803.

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

Giordani, P. and R. Kohn, “Efficient Bayesian inference for multiple change-point and mixture innovation models,” *Journal of Business & Economic Statistics* 26 (2012), 66–77.

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

Green, P. J., “Reversible jump Markov chain Monte Carlo computation and Bayesian model determination,” *Biometrika* 82 (1995), 711–732.

Groen, J. J., G. Kapetanios and S. Price, “Multivariate methods for monitoring structural change,” *Journal of Applied Econometrics* 28 (2013), 250–274.

Hamilton, J. D. and M. T. Owyang, “The propagation of regional recessions,” *Review of Economics and Statistics* 94 (2012), 935–947.

Hong, H., W. Torous and R. Valkanov, “Do industries lead stock markets?,” *Journal of Financial Economics* 83 (2007), 367–396.

Horváth, L. and M. Hušková, “Change-point detection in panel data,” *Journal of Time Series Analysis* 33 (2012), 631–648.

Jagannathan, R. and T. Ma, “Risk reduction in large portfolios: Why imposing the wrong constraints helps,” *Journal of Finance* 58 (2003), 1651–1683.

Kapur, S., “Technological diffusion with social learning,” *Journal of Industrial Economics* 43 (1995), 173–195.

Kass, R. E. and A. E. Raftery, "Bayes factors," *Journal of the American Statistical Association* 90 (1995), 773–795.

Kim, D., "Estimating a common deterministic time trend break in large panels with cross sectional dependence," *Journal of Econometrics* 164 (2011), 310–330.

———, "Common breaks in time trends for large panel data with a factor structure," *The Econometrics Journal* 17 (2014), 301–337.

Koop, G. and S. Potter, "A flexible approach to parametric inference in nonlinear and time varying time series models," *Journal of Econometrics* 159 (2010), 134–150.

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

Nickell, S., "Biases in dynamic models with fixed effects," *Econometrica* 49 (1981), 1417–1426.

Oka, T. and Z. Qu, "Estimating structural changes in regression quantiles," *Journal of Econometrics* 162 (2011), 248–267.

Pástor, L. and R. F. Stambaugh, "The equity premium and structural breaks," *Journal of Finance* 56 (2001), 1207–1239.

Peluso, S., S. Chib and A. Mira, "Semiparametric Multivariate and Multiple Change-Point Modelling," *Unpublished Manuscript* (2016).

Pesaran, M. H., "General diagnostic tests for cross section dependence in panels," *Unpublished Manuscript* (2004).

———, "Estimation and inference in large heterogeneous panels with a multifactor error structure," *Econometrica* 74 (2006), 967–1012.

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

Qian, J. and L. Su, "Shrinkage estimation of common breaks in panel data models via adaptive group fused lasso," *Journal of Econometrics* 191 (2016), 86–109.

Qu, Z. and P. Perron, "Estimating and testing structural changes in multivariate regressions," *Econometrica* 75 (2007), 459–502.

Rapach, D. E., J. K. Strauss and G. Zhou, "Out-of-sample equity premium prediction: Combination forecasts and links to the real economy," *Review of Financial Studies* 23 (2010), 821–862.

———, "International stock return predictability: what is the role of the United States?," *Journal of Finance* 68 (2013), 1633–1662.

Rossi, B., "Advances in forecasting under instability," *Handbook of Economic Forecasting (Elliott, G and Timmermann, A. (Eds.))* Chapter 21 (2013), 1203–1324.

Smith, S. C. and A. Timmermann, "Break Risk," *Unpublished Manuscript* (2017a).

———, "Detecting breaks in real time: A panel forecasting approach," *Unpublished Manuscript* (2017b).

Smith, S. C., A. Timmermann and Y. Zhu, "Variable selection in panel models with breaks," *Unpublished Manuscript* (2017).

Stock, J. H. and M. W. Watson, "Evidence on structural instability in macroeconomic time series relations," *Journal of Business & Economic Statistics* 14 (1996), 11–30.

Wachter, J. A. and M. Warusawitharana, "Predictable returns and asset allocation: Should a skeptical investor time the market?," *Journal of Econometrics* 148 (2009), 162–178.

Wang, J. and E. Zivot, "A Bayesian time series model of multiple structural changes in level, trend, and variance," *Journal of Business & Economic Statistics* 18 (2000), 374–386.

Table 1: Simulation study

	Series 1				Series N			
Parameter	Regime/Break number							
	1	2	3	4	1	2	3	4
	Noncommon breaks							
β	0.973	2.509	2.584	5.029	0.973	2.509	2.584	2.584
	(0.039)	(0.045)	(0.091)	(0.159)	(0.039)	(0.045)	(0.091)	(0.091)
σ	0.987	1.003	2.472	2.408	0.987	1.003	2.472	2.42
	(0.031)	(0.031)	(0.067)	(0.097)	(0.031)	(0.031)	(0.067)	(0.067)
p	0.917	0.913	0.584					
	(0.078)	(0.079)	(0.142)					
	Common breaks (pooled)							
β	0.972	2.501	2.501	3.843	0.972	2.501	2.501	3.843
	(0.049)	(0.048)	(0.121)	(0.126)	(0.049)	(0.048)	(0.121)	(0.126)
σ	0.989	1.011	2.468	2.668	0.989	1.011	2.468	2.668
	(0.034)	(0.066)	(0.081)	(0.085)	(0.034)	(0.066)	(0.081)	(0.085)
	Common breaks (unit-specific)							
β	0.935	2.436	3.266	3.266	0.747	2.234	2.404	2.404
	(0.172)	(0.164)	(0.254)	(0.254)	(0.225)	(0.142)	(0.250)	(0.250)
σ	1.161	1.079	2.421	2.421	1.123	1.116	2.529	2.529
	(0.117)	(0.138)	(0.174)	(0.174)	(0.112)	(0.149)	(0.180)	(0.180)
	Simulated values							
β	1	2.5	2.5	4.75	1	2.5	2.5	2.5
σ	1	1	2.5	2.5	1	1	2.5	2.5
p	1	1	0.6					

Table 1: Simulation study. This table displays the results from estimating the noncommon breaks model developed in this article along with the common breaks models with pooled and unit-specific parameters, respectively, using the hyperparameters and simulated data detailed in Section 5.1 with a cross-section dimension $N = 10$. Estimated posterior means are reported along with standard deviations in brackets below. The true parameter values are presented in the bottom panel. The left panel displays estimates for series 1 which is hit by all three breaks while the right panel displays results for series N which is only hit by the first two breaks.

Table 2: Root mean squared error of estimated break dates

Model	Break number		
	1	2	3
Common (pooled)	0	2	31.686
Common (unit-specific)	0	2	37.202
Noncommon	0.447	1.897	0.894

Table 2: Root mean squared error of break estimates. This table displays the average across the $N = 10$ series of the root mean square error of each of the three breaks estimated from the common break models with pooled and unit-specific parameters, respectively, and the noncommon breaks model.

Table 3: Summary statistics

Country	Mean	St. dev.	Min.	Max.	Autocorrelation	SR
Austria	0.11	5.71	-28.30	27.09	0.21	0.02
Belgium	0.42	4.79	-31.77	23.55	0.14	0.09
Canada	0.34	4.54	-23.18	16.01	0.10	0.07
Denmark	0.41	4.94	-18.92	18.52	0.14	0.08
Finland	0.72	6.60	-27.28	32.21	0.25	0.11
France	0.51	5.71	-22.51	22.39	0.10	0.09
Germany	0.08	5.29	-24.32	19.57	0.08	0.01
Hong Kong	1.05	9.19	-43.88	67.06	0.06	0.11
Italy	0.28	6.75	-20.64	28.61	0.08	0.04
Japan	0.31	5.08	-21.72	17.53	0.11	0.06
Malaysia	0.61	7.76	-34.99	35.89	0.11	0.08
Netherlands	0.65	5.12	-22.76	21.99	0.07	0.13
Singapore	0.51	7.34	-42.11	47.10	0.09	0.07
Spain	0.33	6.01	-26.11	26.15	0.14	0.06
Switzerland	0.48	4.51	-24.94	22.16	0.13	0.11
United Kingdom	0.48	5.54	-27.29	53.22	0.11	0.09
United States	0.53	4.39	-21.97	16.13	0.03	0.12
World	0.45	4.28	-19.37	14.17	0.09	0.11

Table 3: Summary statistics. This table displays summary statistics for monthly excess returns in percent for each of the countries included in our sample and for a worldwide series. The excess return is the return on a broad market index minus the three-month Treasury-bill rate. The Sharpe Ratio (SR) is computed as the mean excess return divided by its standard deviation (St. dev.). All data are sourced from the Global Financial Database.

Table 4: Break dynamics

	dy		tbl		tms		pe	
Break number	p	δ	p	δ	p	δ	p	δ
1	0.796	1.958	0.682	2.026	0.663	1.841	0.703	1.924
	(0.095)	(0.699)	(0.110)	(0.707)	(0.084)	(0.644)	(0.123)	(0.746)
2	0.828	3.761	0.873	2.981	0.779	3.560	0.682	2.870
	(0.111)	(0.938)	(0.157)	(0.840)	(0.135)	(0.884)	(0.122)	(0.781)
3	0.524	1.977	0.691	1.587	0.611	2.314	0.544	1.855
	(0.142)	(0.799)	(0.092)	(0.649)	(0.188)	(0.509)	(0.151)	(0.758)
4	0.958	2.296	0.907	2.100	0.898	2.203	0.888	2.470
	(0.064)	(1.291)	(0.095)	(1.190)	(0.062)	(1.111)	(0.050)	(1.291)

Table 4: Break dynamics. This table displays the posterior mean and standard deviation (in brackets below) of the pervasion parameter p and the diffusion parameters δ for each of the breaks detected by the predictive regression of country-level excess stock returns on an intercept and a lagged predictive variable with noncommon breaks. All parameters are pooled. The predictive variables include the dividend yield (dy), the Treasury-bill rate (tbl), the term spread (tms), and the price-earnings ratio (pe).

Table 5: Group structure

Group	Countries						
1	Hong Kong	Malaysia	Singapore				
2	Canada	France	Germany	Italy	Japan	U.K.	U.S.
3	Austria	Belgium	Denmark	Finland	Netherlands	Spain	Switzerland

Table 5: Group structure. This table displays the countries that belong to each group that is inferred from the pervasiveness of breaks estimated by the noncommon breaks model when regressing country-level excess stock returns on an intercept and their corresponding dividend yields using the full sample. Regression coefficients are pooled (and thus identical) across all countries within, but differ between, regime-specific groups.

Table 6: Statistical significance from pairwise MSE comparisons of forecasts from the model with noncommon breaks versus forecasts from a range of benchmarks

Predictor	DM				CW			
	t <-1.64	-1.64< t <0	0< t <1.64	t >1.64	t <-1.64	-1.64< t <0	0< t <1.64	t >1.64
				Time series linear model				
dy	0	1	1	15†	0	1	1	15†
tbl	0	2	1	14†	0	2	1	14†
tms	0	1	1	13†	0	1	0	14†
pe	1	0	2	10†	1	0	1	11†
				Historical average				
dy	0	2	2	13†	0	2	1	14†
tbl	0	1	1	15†	0	1	1	15†
tms	0	1	1	13†	0	1	0	14†
pe	1	0	2	10†	1	0	1	11†
				Pooled constant panel model				
dy	0	0	3	14†	0	0	2	15†
tbl	0	1	1	15†	0	1	0	16†
tms	0	1	1	13†	0	1	1	13†
pe	1	0	0	12†	1	0	0	12†
				Pooled panel common break model				
dy	0	1	1	15†	0	1	1	15†
tbl	0	1	2	14†	0	1	2	14†
tms	0	1	1	13†	0	1	1	13†
pe	0	1	1	11†	0	1	1	11†
				Heterogeneous panel common break model				
dy	0	0	1	16†	0	0	1	16†
tbl	0	1	2	14†	0	1	1	15†
tms	0	1	1	13†	0	1	0	14†
pe	1	0	1	11†	1	0	0	12†

Table 6: Statistical significance of pairwise forecast comparisons. This table evaluates the statistical significance of the improved forecasting power of the model that allows for noncommon breaks relative to a range of benchmarks. The benchmark models include the historical average, the linear time series model, the panel common break models with pooled and unit-specific parameters, and the constant pooled panel model. Aside from the historical average, each model includes an intercept and a single predictor. The predictor is either the dividend yield (dy), T-bill rate (tbl), term spread (tms), or price-earnings ratio (pe). Significance is evaluated using the Diebold-Mariano (DM) and Clark-West (CW) tests. For each procedure the table displays the number of countries for which our method produces significantly worse, insignificantly worse, insignificantly better, and significantly better forecasts at the 5% level. † indicates the bin in which lies the *t*-statistic for the U.S.. All results assume a ten-year warm-up estimation window with out-of-sample forecasts generated recursively for the sample period January 1983 through November 2017.

Table 7: Country allocation

Country	Hist avg	OLS	No Brk	Pbrk	Pbrk (hetero)	Noncommon
Austria	0.01	0.12	0.00	0.02	0.10	0.01
Belgium	0.00	0.04	0.00	0.01	0.02	0.01
Canada	0.00	0.00	0.01	0.00	0.00	0.11
Denmark	0.00	0.00	0.02	0.01	0.00	0.00
Finland	0.16	0.12	0.22	0.06	0.17	0.04
France	0.08	0.05	0.11	0.09	0.06	0.08
Germany	0.23	0.19	0.20	0.13	0.09	0.05
Hong Kong	0.06	0.02	0.02	0.12	0.10	0.18
Italy	0.02	0.00	0.01	0.07	0.04	0.03
Japan	0.10	0.07	0.04	0.10	0.03	0.06
Malaysia	0.01	0.00	0.00	0.03	0.12	0.17
Netherlands	0.00	0.00	0.01	0.00	0.06	0.00
Singapore	0.04	0.12	0.08	0.07	0.09	0.16
Spain	0.00	0.00	0.01	0.03	0.02	0.00
Switzerland	0.00	0.00	0.00	0.01	0.03	0.00
U.K.	0.17	0.21	0.13	0.16	0.01	0.01
U.S.	0.12	0.06	0.13	0.09	0.05	0.08

Table 7: Allocations between countries. This table reports the average weight of the risky portfolio allocated to the risky asset (aggregate stock market) of each of the 17 countries across the out-of-sample period when forecasting with the competing models using the dividend yield as the predictive variable. ‘Hist avg’ denotes the country prevailing mean, ‘OLS’ denotes the time series linear model, ‘No brk’ denotes the pooled panel model without breaks, ‘Pbrk’ denotes the pooled panel common break model, ‘Pbrk (hetero)’ denotes the heterogeneous panel model with common breaks, and ‘Noncommon’ denotes the panel model with noncommon breaks.

Table 8: Utility gains

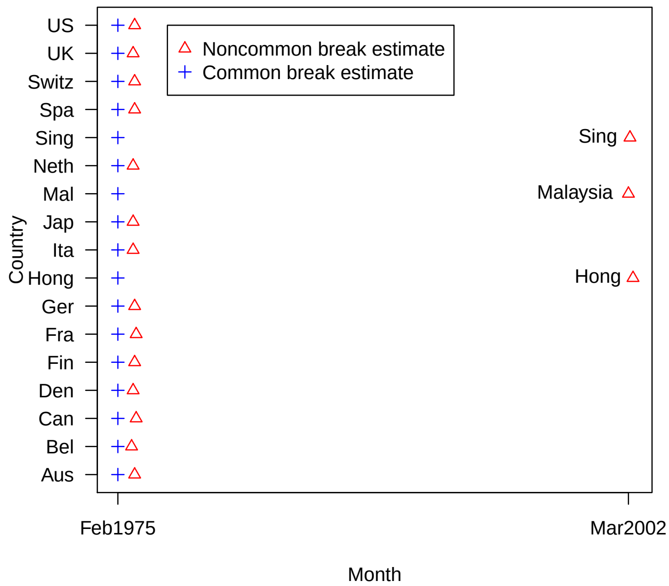
Predictor	Hist avg	OLS	No Brk	Pbrk	Pbrk (hetero)
dy	2.46	2.62	2.01	3.23	2.03
tbl	1.98	3.20	2.54	2.88	3.24
tms	2.41	1.89	2.85	2.46	2.59
pe	1.90	1.95	3.09	2.68	2.08

Table 8: Utility gains. This table reports the out-of-sample utility gain for a mean-variance investor with a risk aversion coefficient of three who at each period allocates his wealth between a risk-free asset (the U.S. T-bill) and an optimal risky portfolio that is constructed from the risky assets of the 17 countries. We report the utility gain measured relative to each of the five benchmark models, namely, the prevailing mean (hist avg), the linear time series model (OLS), the pooled panel model with no breaks (No Brk), and the panel models with common breaks and either pooled (Pbrk) or unit-specific (Pbrk (hetero)) parameters. The utility gain is computed across the out-of-sample period. Results are presented for the four predictors we consider: the dividend yield (dy), the T-bill rate (tbl), the term spread (tms), and the price-earnings ratio (pe).

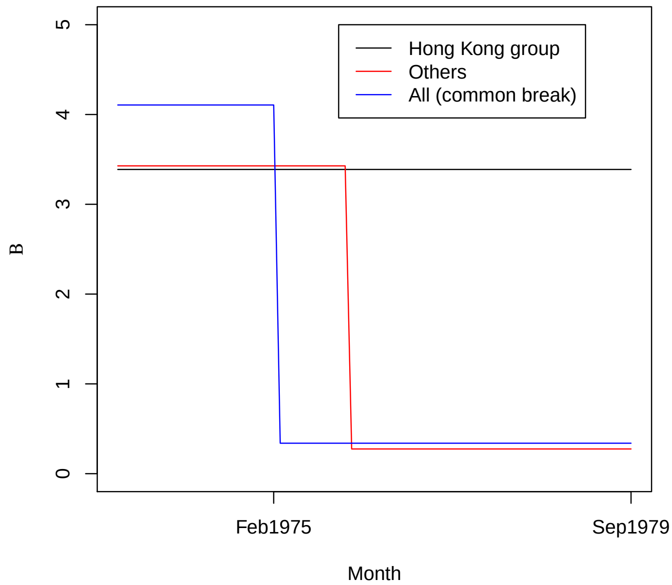
Table 9: Sensitivity of forecasting performance to choice of priors

Hyp. value	K	DM				CW			
		t <-1.64	-1.64< t <0	0< t <1.64	t >1.64	t <-1.64	-1.64< t <0	0< t <1.64	t >1.64
					Time series linear model				
c=120	5	0	1	1	$15^{\dagger}$	0	1	1	$15^{\dagger}$
e=5	4	0	1	1	$15^{\dagger}$	0	1	1	$15^{\dagger}$
g=10	4	0	1	1	$15^{\dagger}$	0	1	1	$15^{\dagger}$
b = 6	4	0	1	1	$15^{\dagger}$	0	1	1	$15^{\dagger}$
$\sigma_{\eta}^2 = 0.01$	4	0	1	1	$15^{\dagger}$	0	1	1	$15^{\dagger}$
$\sigma_{\eta}^2 = 1$	4	0	1	2	$14^{\dagger}$	0	1	2	$14^{\dagger}$
					Historical average				
c=120		0	2	2	$13^{\dagger}$	0	2	1	$14^{\dagger}$
e=5		0	2	2	$13^{\dagger}$	0	2	2	$13^{\dagger}$
g=10		0	2	2	$13^{\dagger}$	0	2	1	$14^{\dagger}$
b = 6		0	2	2	$13^{\dagger}$	0	2	1	$14^{\dagger}$
$\sigma_{\eta}^2 = 0.01$		0	2	1	$14^{\dagger}$	0	2	0	$15^{\dagger}$
$\sigma_{\eta}^2 = 1$		0	2	2	$13^{\dagger}$	0	2	2	$13^{\dagger}$
					Pooled constant panel model				
c=120		0	1	3	$13^{\dagger}$	0	1	2	$14^{\dagger}$
e=5		0	0	3	$14^{\dagger}$	0	0	3	$14^{\dagger}$
g=10		0	0	3	$14^{\dagger}$	0	0	2	$15^{\dagger}$
b = 6		0	0	3	$14^{\dagger}$	0	0	2	$15^{\dagger}$
$\sigma_{\eta}^2 = 0.01$		0	0	3	$14^{\dagger}$	0	0	2	$15^{\dagger}$
$\sigma_{\eta}^2 = 1$		0	0	3	$14^{\dagger}$	0	0	3	$14^{\dagger}$
					Pooled panel common break model				
c=120		0	1	1	$15^{\dagger}$	0	1	1	$15^{\dagger}$
e=5		0	1	1	$15^{\dagger}$	0	1	1	$15^{\dagger}$
g=10		0	1	1	$15^{\dagger}$	0	1	1	$15^{\dagger}$
b = 6		0	1	1	$15^{\dagger}$	0	1	1	$15^{\dagger}$
$\sigma_{\eta}^2 = 0.01$		0	1	2	$14^{\dagger}$	0	1	2	$14^{\dagger}$
$\sigma_{\eta}^2 = 1$		0	1	$1^{\dagger}$	15	0	1	$1^{\dagger}$	15
					Heterogeneous panel common break model				
c=120		0	1	1	$15^{\dagger}$	0	1	1	$15^{\dagger}$
e=5		0	1	1	$15^{\dagger}$	0	1	1	$15^{\dagger}$
g=10		0	0	2	$15^{\dagger}$	0	0	1	$16^{\dagger}$
b = 6		0	1	1	$15^{\dagger}$	0	1	1	$15^{\dagger}$
$\sigma_{\eta}^2 = 0.01$		0	0	2	$15^{\dagger}$	0	0	1	$16^{\dagger}$
$\sigma_{\eta}^2 = 1$		0	1	2	$14^{\dagger}$	0	1	1	$15^{\dagger}$

Table 9: Sensitivity of forecasting results to priors. This table reports results from forecasting country-level excess stock returns using the model with noncommon breaks relative to the benchmark models that include the historical average, the linear time series model, the panel common break models with pooled and unit-specific parameters, and the pooled panel model with no breaks. Aside from the historical average, each model includes an intercept and the dividend yield as the predictive variable. Adjusting one hyperparameter at a time we substitute the value in the table for its value in Section 6.4. The remaining hyperparameters are unchanged. The five panels display the statistical significance of whether our model produces more or less accurate forecasts relative to the five benchmarks. Significance is evaluated using the Diebold-Mariano (DM) and Clark-West tests (CW). The table displays the number of countries for which our method produces significantly worse, insignificantly worse, insignificantly better, and significantly better forecasts at the 5% level. † indicates the bin in which falls the U.S.. The first column displays the estimated posterior mode of the number of breaks K from the model with noncommon breaks using the full sample.

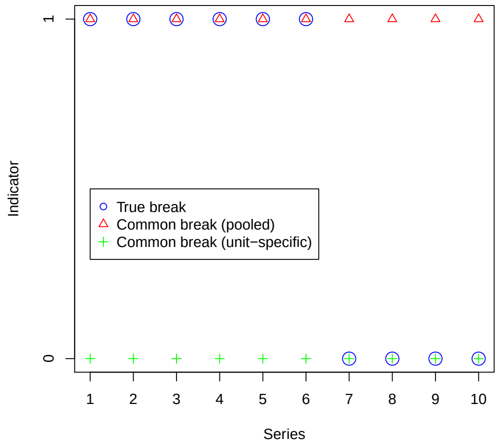


(a) Oil price shock (1975)

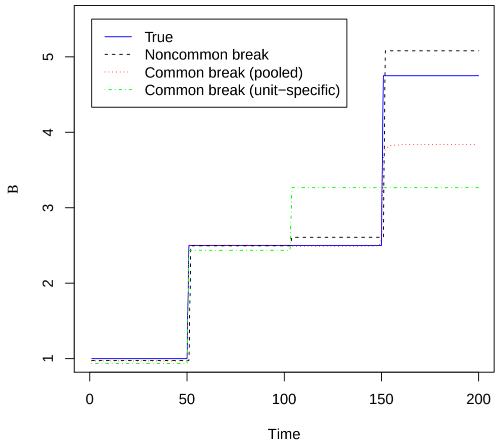


(b) Dividend yield slope

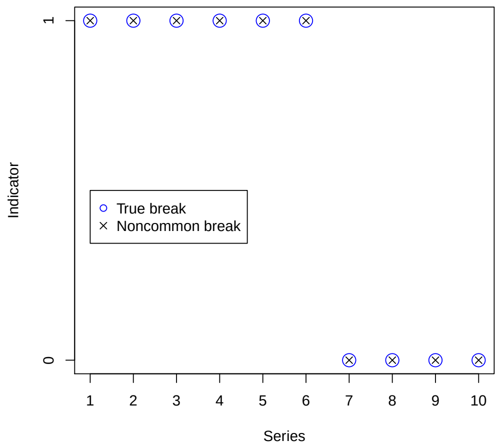
Figure 1: The top left window displays the first break. The blue crosses (red triangles) mark the posterior mode estimate of the first break for each country from the pooled panel model with common (noncommon) breaks that regresses excess stock returns on an intercept and lagged dividend yields. The bottom window graphs the path of the slope coefficient. The blue line graphs the path for all countries from the common breaks model. Estimates from the noncommon breaks model are represented by the black line for Hong Kong, Malaysia and Singapore, and the red line for the 14 remaining countries.



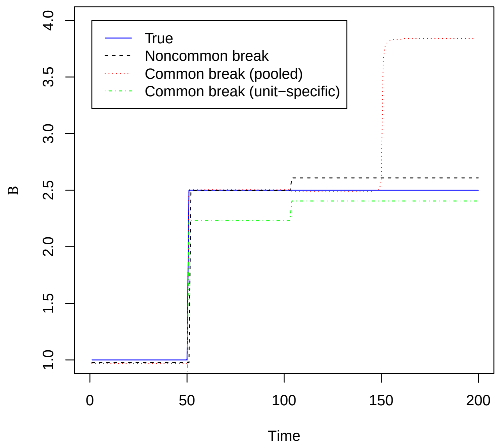
(a) Break 1



(b) β_i for $i \in N_3$

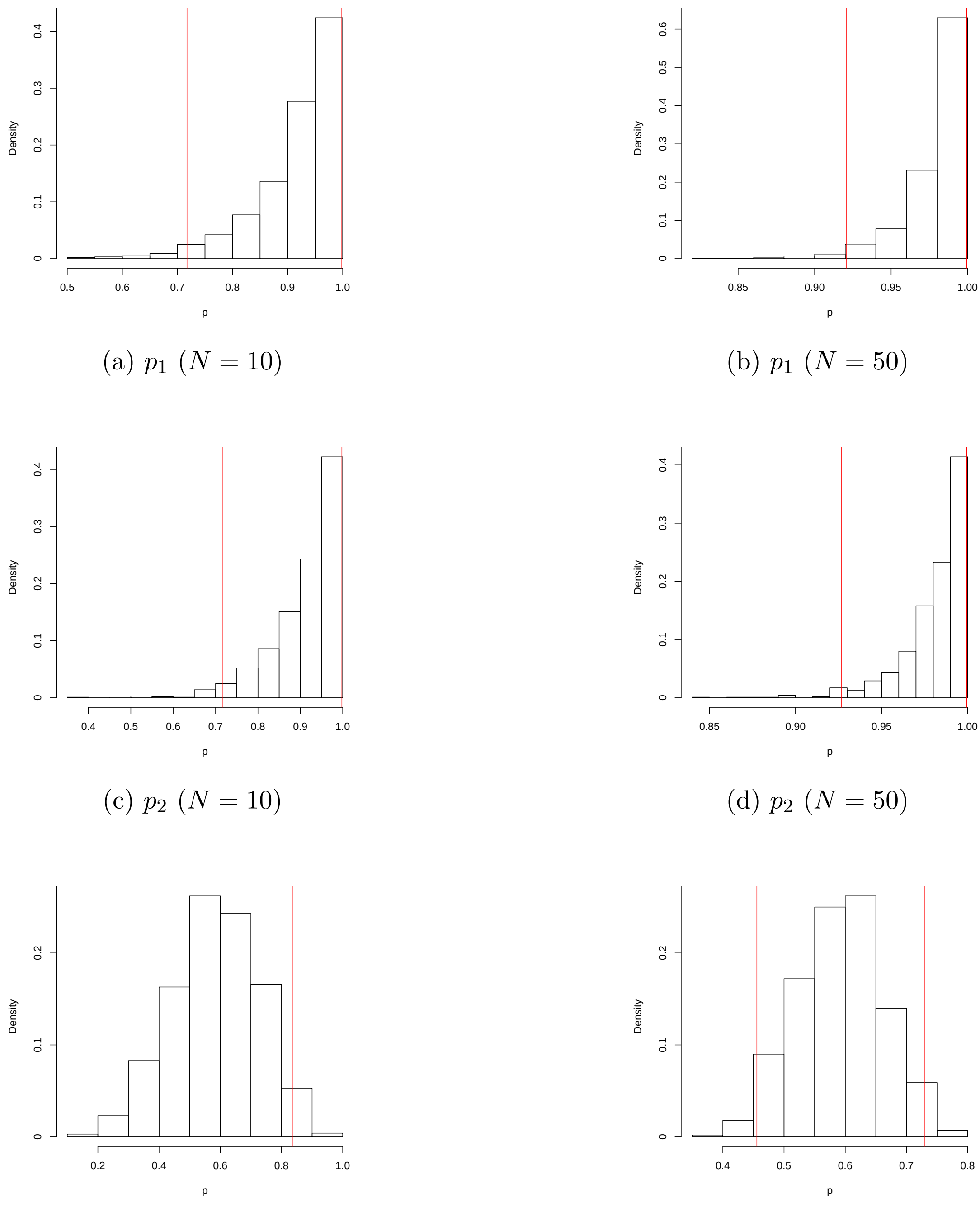


(c) Break 2



(d) β_i for $i \notin N_3$

Figure 2: The left panel of this figure displays the true break dates (blue circles) and the estimated posterior mode of whether each series is hit (Indicator = 1) by the third break when the cross-sectional dimension is $N = 10$. The top left window displays break estimates from the common break panel with pooled (red triangles) and unit-specific (green crosses) coefficients. The bottom left window displays break estimates from the noncommon break model (black crosses). The top right window graphs the evolution of the estimated slope coefficient from the noncommon breaks model (black dashed line), the common break panel model with unit-specific (green dotted line) and pooled (red dotted line) coefficients, and the true path (blue solid line) for those series hit by the third break $i \in N_3$. The bottom right window graphs the slopes for series not hit by the third break $i \notin N_3$.



(a) p_1 $(N = 10)$ (b) p_1 $(N = 50)$

(c) p_2 $(N = 10)$ (d) p_2 $(N = 50)$

(e) p_3 $(N = 10)$ (f) p_3 $(N = 50)$

Figure 3: This figure displays the estimated posterior distribution of the pervasion parameter p when the cross-section dimension is $N = 10$ (left panel) and $N = 50$ (right panel), respectively for the three breaks. The vertical red lines mark the top and bottom 2.5% of the distribution.

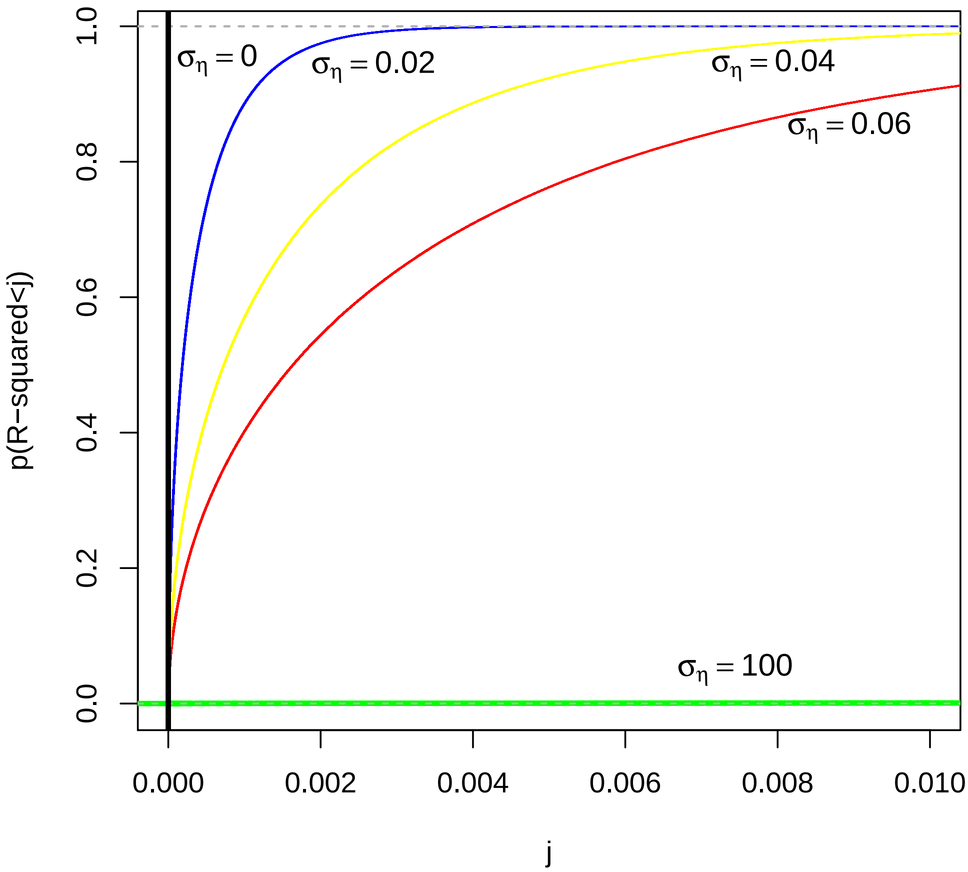
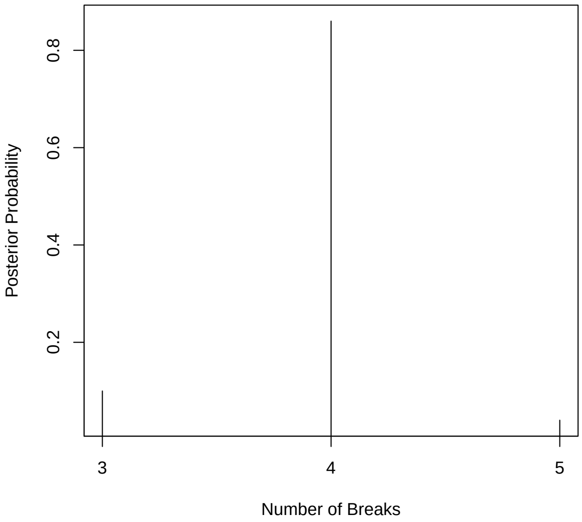
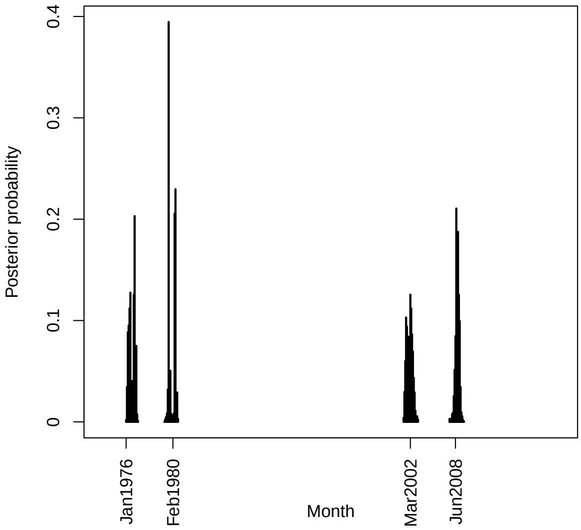


Figure 4: This Figure displays, for varying amounts of scepticism surrounding the degree of predictability prevalent in the market, the prior probability that the R-squared value from a predictive regression falls below a certain value j that ranges from 0 to 0.01. A zero value of σ_η suggests investors have a dogmatic belief that returns are unpredictable, an infinite value of $\sigma_\eta = 0$ suggest investors believe any degree of return predictability is equally likely and positive but finite values of σ_η suggests investors are sceptical about the existence of return predictability but do not altogether rule it out.

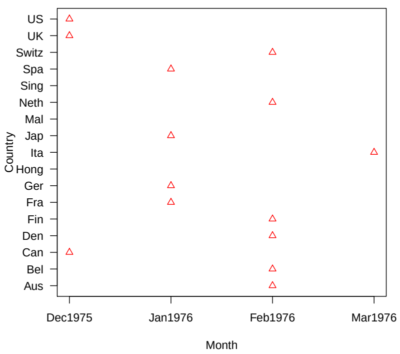


(a) Posterior model probabilities

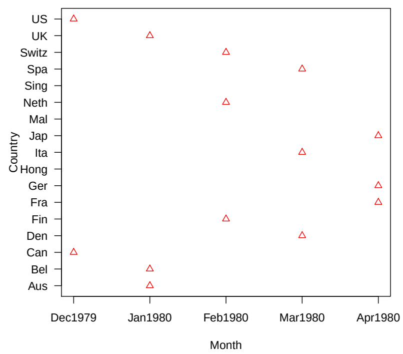


(b) Posterior break locations

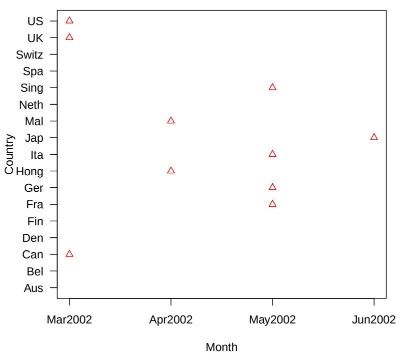
Figure 5: This figure displays the posterior model probabilities (top window) and posterior break locations (bottom window) estimated from the pooled panel predictive regression with noncommon breaks when regressing country-level excess stock returns on an intercept and country-level lagged dividend yields.



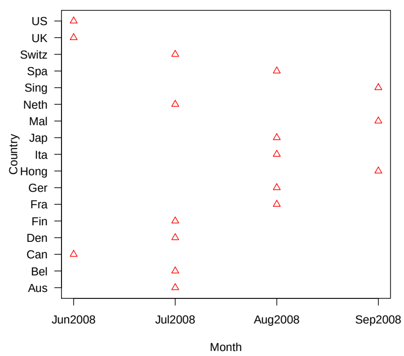
(a) 1st oil price shock



(b) 2nd oil price shock

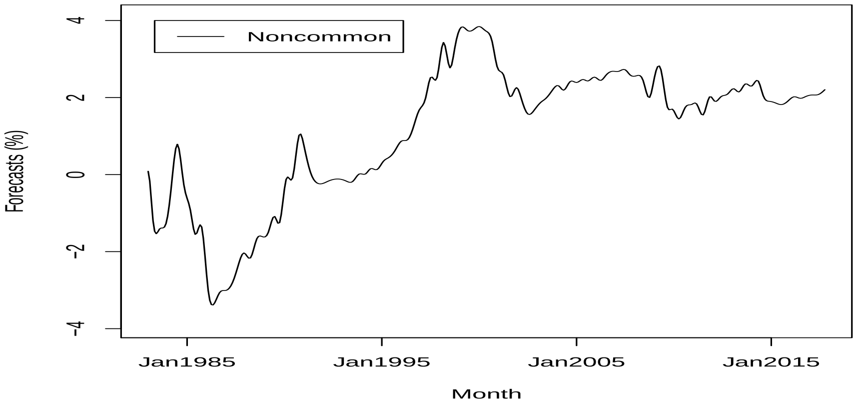


(c) Dotcom bubble bursting

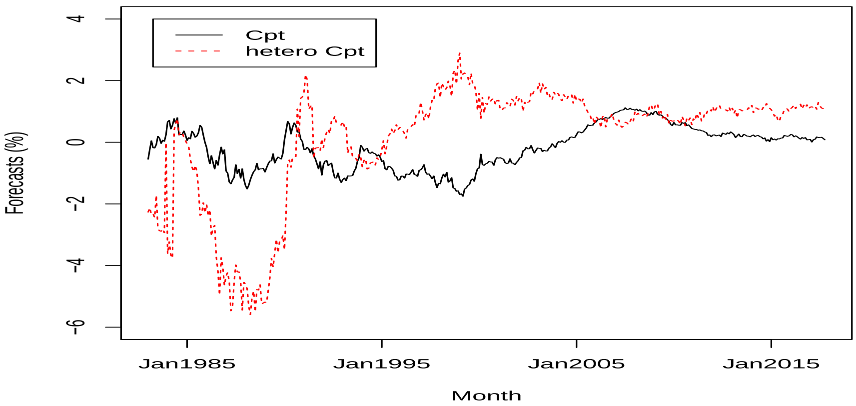


(d) Global financial crisis

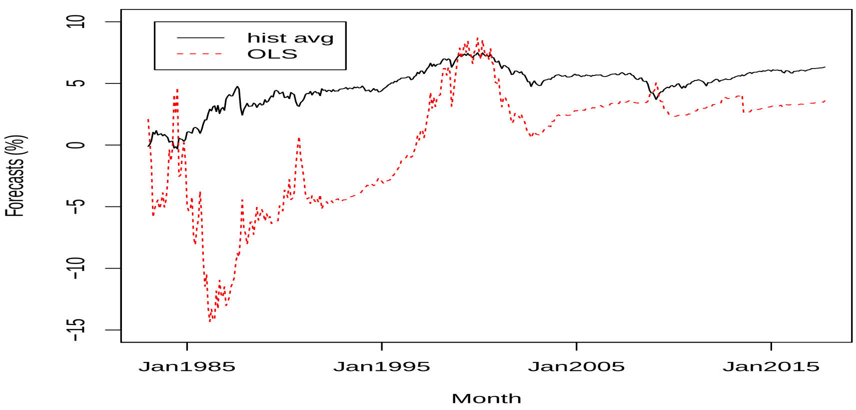
Figure 6: This figure displays the posterior modes of each of the breaks estimated from the predictive panel regression model with noncommon breaks using country-level dividend yields as the predictive variable. The red triangles mark the time at which each country is hit by the break in question while no red triangle implies the corresponding country is not hit. The top left, top right, bottom left, and bottom right windows correspond to the first oil price shock, second oil price shock, bursting of the dotcom bubble, and the global financial crisis.



(a) Noncommon break model

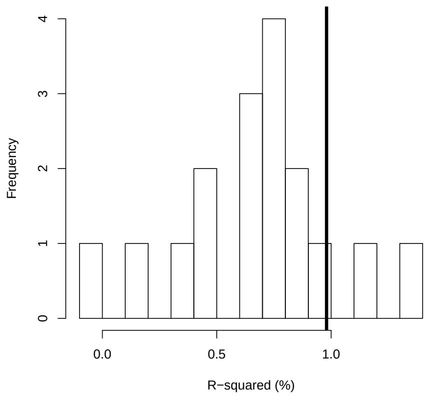


(b) Panel break models

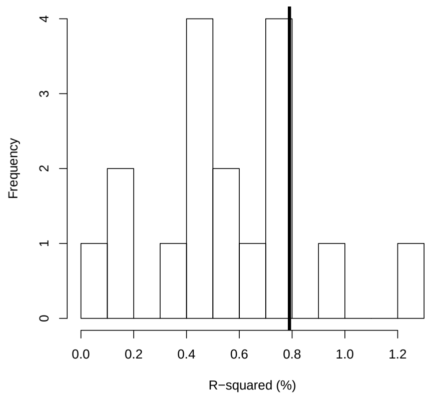


(c) Time series models

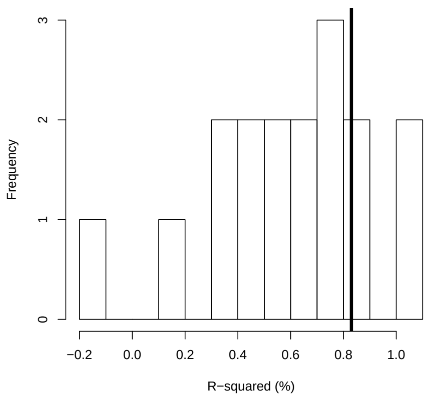
Figure 7: This figure graphs the forecasts of U.S. excess stock returns generated from the model with noncommon breaks (top panel). The middle panel displays forecasts generated from the models with common breaks and either pooled (black solid line) or unit-specific parameters (red dashed line). The bottom panel displays the forecasts generated from the linear time series model (red dashed line) and the historical average (black solid line). Aside from the historical average, each model generates forecasts using the dividend yield as the predictive variable alongside an intercept.



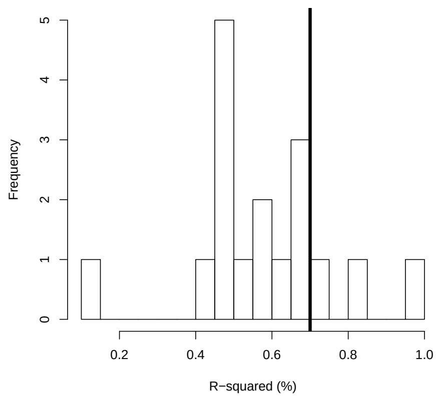
(a) Time series linear model



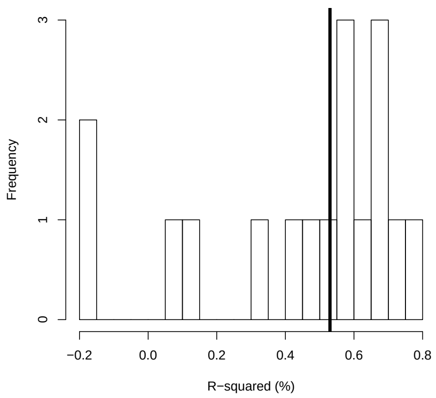
(b) Pooled Panel No Breaks



(c) Pooled Panel Breaks

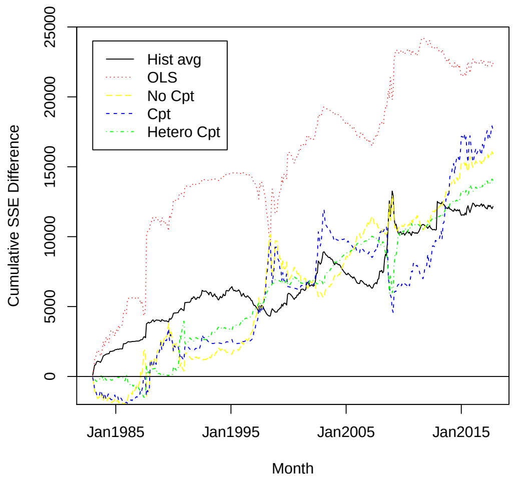


(d) Heterogeneous Panel Breaks

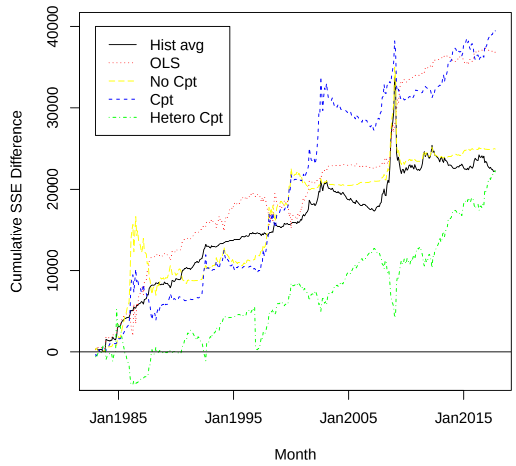


(e) Historical Average

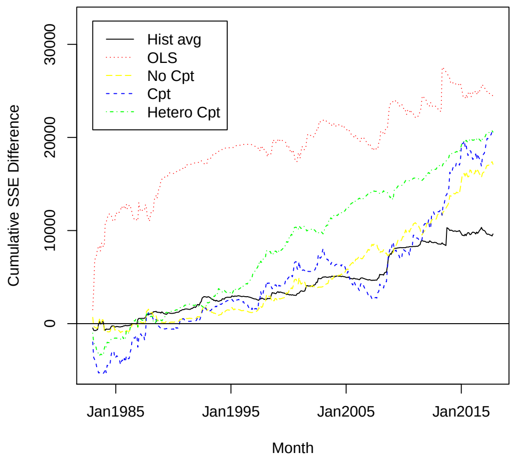
Figure 8: This figure displays the R^2_{OoS} values obtained from comparing the forecasting performance of the model that allows for noncommon breaks with the benchmark model labeled in the subcaption for each of the seventeen countries (the U.S. is marked by the black vertical line). Except for the historical average, all models include an intercept and the dividend yield as the predictive variable.



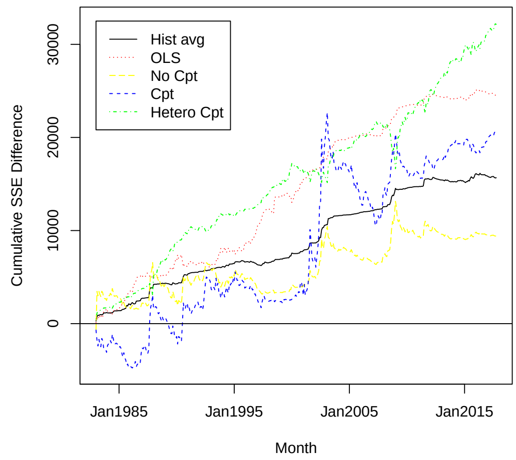
(a) U.S.



(b) Italy

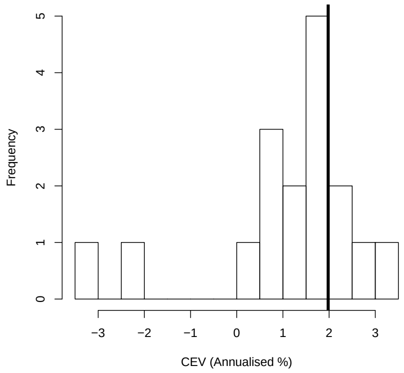


(c) U.K.

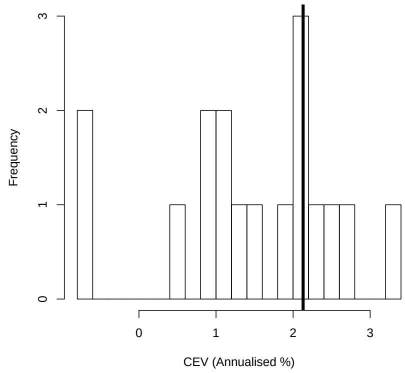


(d) Germany

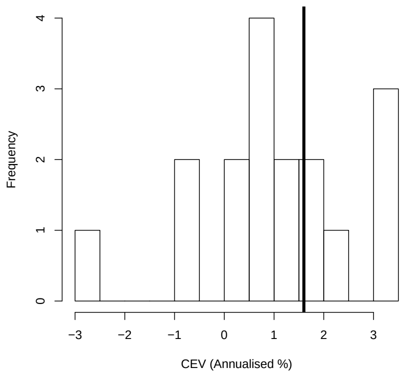
Figure 9: This figure graphs the cumulative difference in the sum of squared forecast errors for U.S., Italy, U.K., and Germany obtained from the noncommon breaks model relative to each of the benchmarks. The competing models are the historical average, the time series linear model, the panel models with common breaks and pooled or unit-specific parameters, and the pooled panel model with no breaks. Except for the historical average, all models include an intercept and the dividend yield as the predictive variable.



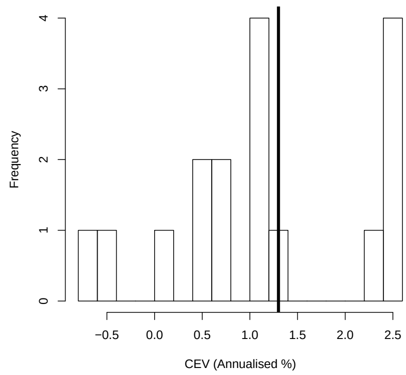
(a) Time series linear model



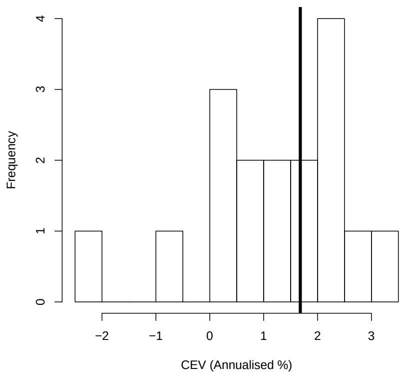
(b) Pooled Panel No Breaks



(c) Pooled Panel Breaks



(d) Heterogeneous Panel Breaks



(e) Historical Average

Figure 10: This Figure displays the out-of-sample utility gain to a mean-variance investor with a risk aversion coefficient of three who at each period allocates his wealth between the risky asset and the risk-free rate of the given country. Utility gains are reported as annualised percentages obtained when comparing the forecasting performance of the noncommon breaks model with the benchmark model labeled in the subcaption for each of the 17 countries (the U.S. is marked by the black vertical line). Except for the historical average, all models include an intercept and the dividend yield as the predictive variable.