Break Risk

Simon C. Smith
USC Dornsife INET

Allan Timmermann
University of California San Diego

AFA 2019 Annual Meeting

Atlanta

January 6, 2019

Evidence of model instability

- Stock return forecasts are plagued by model instability
  - Pastor and Stambaugh (2001): breaks in equity premium
  - Viceira (1997), Lettau and Van Nieuwerburgh (2008), and Pettenuzzo and Timmermann (2011): breaks in dividend yield and/or the relation between stock returns and the lagged dividend yield
  - Rapach and Wohar (2006) and Paye and Timmermann (2006): breaks in the relation between returns and various predictors for the US and internationally

Potential causes of model instability

- Predictability patterns may ‘self destruct’ as investors attempt to exploit them
  - Schwert (2003), Green et al (2011), and McLean and Pontiff (2016): abnormal returns disappear or are greatly reduced after becoming public knowledge, due to competitive pressure
- Technological disruption, institutional, regulation, and policy changes
  - Firms may shift from paying dividends to share repurchases if taxes on dividends rise

Difficulties in exploiting model instability for prediction

- Lettau and Van Nieuwerburgh (2008): shifts in dividend yield cannot be exploited for two reasons
  1. Difficult to detect and determine location of breaks in real time (timing)
  2. Difficult to estimate magnitude of break, especially if break is recent (tracking)
- Using full-sample estimates will bias parameter estimates if shifts have occurred
- Using a shorter window of returns (after a break) leads to large estimation errors and inaccurate forecasts

Exploiting information in the cross-section

- Lettau and Van Nieuwerburgh (2008) use a univariate time-series approach to detect breaks
  - Time series break methods typically only detect the largest breaks and with considerable delay
- Panel data sets widely used in finance - exploiting the information in the cross-section increases our ability to detect breaks
  - also helps forecast aggregate returns using disaggregate components (Ferreira and Santa-Clara (2011), Kelly and Pruitt (2013))
- If breaks are relatively pervasive and their **timing** is relatively homogeneous, a panel break procedure may offer more power compared with time series approach

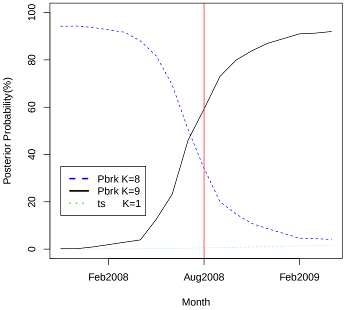
Estimating magnitude of parameter shifts

- A Bayesian approach enables a prior to be specified that suggests large shifts in the parameters are unlikely
  - Pastor and Stambaugh (2001) specify a prior that suggests large shifts in the equity premium are unlikely
- Following Wachter and Warusawitharana (2009) we specify a prior which suggests investors are sceptical about the existence of return predictability - directly maps to prior over R^2

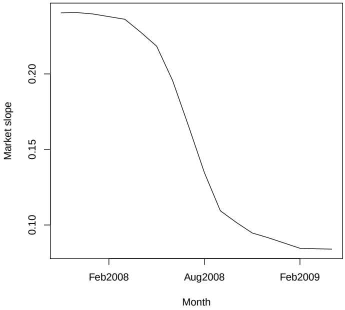
Forecasting industry portfolio returns

- The ability of predictors to forecast the aggregate stock portfolio should carry over to industry portfolios if markets are relatively efficient - information dissemination
  - if a variable ceases to predict returns on the market portfolio it is likely to stop predicting returns on industry portfolios
  - forecasting industry portfolio returns from aggregate predictors may allow us to better estimate the timing and magnitude of breaks

Onset of Financial Crisis



(a) Posterior number of breaks



(b) Slope on d/p

Models

Time series breakpoint model

- K_i unit-specific structural breaks split the sample into $K_i + 1$ distinct regimes with changepoints $\tau_i = (\tau_{i1}, \ldots, \tau_{iK_i})$

$$\tilde{r}_{it} = \mu_{ik} + \tilde{X}_{t-1}\beta_{ik} + \epsilon_{it}, \qquad k = 1, \ldots, K_i + 1, \qquad t = \tau_{k_i - 1} + 1, \ldots, \tau_{k_i}$$

- $\tilde{r}_{it}$: dependent variable for the ith series
- μ_{ik}: intercept in the k_ith regime
- $\tilde{X}_{t-1}$: aggregate predictor at time $t - 1$
- β_{ik}: slope coefficient in the k_ith regime
- $\epsilon_{it} \sim N(0, \sigma^2_{ik})$

Correlated effects

- Industry portfolio returns likely to be highly correlated

1. Estimate correlation matrix in each regime?
   - Many parameters
   - Computationally infeasible
2. Allow for latent common factor error structure (Pesaran, 2006) ✓

Accounting for Dependencies: common correlated effects

- Following Pesaran (2006), assume correlations are induced by a common factor f_t

$$
\begin{aligned}
&\tilde{r}_{it} = \beta_{it}' \tilde{X}_{t-1} + \tilde{\epsilon}_{it}, \qquad i = 1, \ldots, N, \qquad t = 2, \ldots, T, \\
&\tilde{\epsilon}_{it} = \gamma_{it} f_t + \nu_{it}.
\end{aligned}
$$

- ν_{it}: idiosyncratic errors not correlated
- If f_t is observed simply add it to the regression; if it's unobserved use cross-sectional average of dependent variable as a proxy
- Effectively, we account for cross-sectional dependencies in returns by cross-sectionally "de-meaning" (pre-filtering) the data
- Baltagi et al (2016) shows that Pesaran (2006)'s CCE approach remains asymptotically valid in presence of breaks if they are common

Common break models

- Pool information by assuming that break point locations are the same across cross-sectional units
  - common sources of breaks (e.g., Global Financial Crisis)
- Maintain cross-sectional heterogeneity in parameters (for $i = 1, ..., N$)

$$r_{it} = \mu_{ik} + X_{t-1}\beta_{ik} + \epsilon_{it}, \qquad t = \tau_{k-1} + 1, \ldots, \tau_k, \qquad k = 1, \ldots, K+1$$

- Assuming homogeneous parameters and break dates, we get a model

$$r_{it} = \mu_k + X_t\beta_k + \epsilon_{it}, \qquad t = \tau_{k-1} + 1, \ldots, \tau_k, \qquad k = 1, \ldots, K+1$$

Priors on regression coefficients

- Prior on σ^2_{ik} is inverse gamma

$$p(\sigma^2_{ik}) = \frac{b^a}{\Gamma(a)} {\sigma^2_{ik}}^{-(a+1)} exp\left(-\frac{b}{\sigma^2_{ik}}\right)$$

- Pastor and Stambaugh (1999) suggest ruling out implausibly high SRs
- Prior on intercept μ_{ik} is Gaussian conditional on error variance

$$p(\mu_{ik}) = N(0, \sigma^2_\mu \sigma^2_{ik})$$

- Choose moderate $\sigma^2_\mu = 5\%$ following Pastor and Stambaugh (1999)

Prior beliefs about degree of predictability

- High variance of the predictor σ_x^2 might lower the prior on β_i whereas a large residual variance σ_{ik}^2 might increase it (Wachter and Warusawitharana, 2009)
- Place prior over this 'normalised beta' $\eta_{ik} = \beta_{ik} \frac{\sigma_x}{\sigma_{ik}}$

$$\eta_{ik} \sim N(0, \sigma_\eta^2)$$

- Equivalent to placing the following prior on β_{ik}

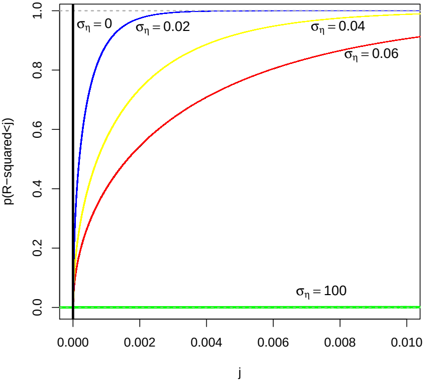
$$p(\beta_{ik}) \sim N\left(0, \frac{\sigma_\eta^2}{\sigma_x^2} \sigma_{ik}^2\right)$$

- Compute σ_x^2 as the empirical variance of the predictor variable over the full sample available at the time the recursive forecast is made
- σ_η controls prior degree of predictability

Prior beliefs about degree of predictability

- This prior implies that no risky asset can have an R^2 that is too large

$$R_i^2 = \frac{\beta_i^2 \sigma_x^2}{\beta_i^2 \sigma_x^2 + \sigma_{ik}^2} = \frac{\eta_i^2}{\eta_i^2 + 1}.$$



Posterior

Proposition 1. Assuming inverse gamma priors on the error-term variances, σ^2, and Gaussian priors on the regression coefficients, β, conditional on σ^2, the posterior distribution of the heterogeneous panel model with breaks after marginalising the parameters takes the form

$$p(\boldsymbol{r} \mid \boldsymbol{X}, \tau) = \prod_{i=1}^{N} \prod_{k=1}^{K+1} (2\pi)^{-l_k/2} \frac{b^a}{\Gamma(a)} \frac{\Gamma(\tilde{a}_k)}{\tilde{b}_{ik}^{\tilde{a}_k}} |\Sigma_k|^{1/2} \mid V_\beta \mid^{-1/2}$$

- where

$$\Sigma_k^{-1} = V_\beta^{-1} + X_k X_k'$$
$$\mu_{ik} = \Sigma_k X_k r_{ik}$$
$$\tilde{a}_k = \tilde{a}_{ik} = a + (l_k)/2$$
$$\tilde{b}_{ik} = \frac{1}{2}\left(2b + r_{ik}' r_{ik} - \mu_{ik}' \Sigma_k^{-1} \mu_{ik}\right)$$

Estimation

Estimation of panel breaks model

- Simulate the changepoint vector τ in two steps
  - global movement: attempt to add or remove a changepoint on each sweep of the MCMC run
  - perturb each changepoint locally by a random-walk Metropolis-Hastings step
- Sample the parameters from their full conditional distributions

Birth move

- With equal probability a birth move is entered
- This move attempts to increase K to $K + 1$ and introduce a new changepoint sampled uniformly from the time series
  - $\tau_{k^*} \sim U[1, T]$
- Split an existing regime into two new ones and compute parameters for the new regimes
- Accept the birth move with a probability $min(1, \alpha)$, where
  - $\alpha = \frac{p(\boldsymbol{r}|\boldsymbol{X},\tau^*)}{p(\boldsymbol{r}|\boldsymbol{X},\tau)} \times \frac{p(\tau^*)}{p(\tau)} \times \frac{T}{K+1} \times \frac{2}{2}$
  - formula for α is given in the paper

Death move

- With equal probability a death move is entered
- This move attempts to decrease K to $K-1$ sampling one of the existing changepoints uniformly
  - $\tau_{k^c} \sim U[\tau_1, \tau_K]$
- Merge two existing regimes into one new one and compute hyperparameters for the new regime
- Accept the death move with probability $min(1, \alpha)$, where
  - $\alpha = \frac{p(\boldsymbol{r}|\boldsymbol{X},\tau^*)}{p(\boldsymbol{r}|\boldsymbol{X},\tau)} \times \frac{p(\tau^*)}{p(\tau)} \times \frac{K}{T} \times \frac{2}{2}$

Estimating changepoint locations

- RW-MH step provides local adjustment
- For $k \in (1, K)$ each changepoint τ_k is perturbed by a discrete number u that is sampled uniformly from the interval $[-s, s]$
- If perturbation $= 0$ the proposal is immediately rejected
- Otherwise compute FC hyperparameters
- Accept with probability $min(1, \alpha)$

$$\alpha = \frac{p(\tau^*)}{p(\tau)} \frac{p(r \mid X, \tau^*, \theta)}{p(r \mid X, \tau, \theta)}$$

Estimating the number of breaks

- Common to use the Bayes factor between two competing models to determine the number of breaks

- Compare a model with $K = 2$ breaks to a model with a single break ($K = 1$) through the ratio

$$BF_{21} = \frac{P(K = 2 \mid r)/P(K = 1 \mid r)}{P(K = 2)/P(K = 1)} = \frac{p(r \mid K = 2)}{p(r \mid K = 1)}$$

- Assuming uniform prior model probabilities, the marginal likelihoods are proportional to the posterior model probabilities

- Chib (1998)'s widely used algorithm fixes the number of breaks in advance which leads to a nonuniform prior distribution on the changepoint locations

Estimating the number of breaks: RJ Approach

- Reversible jump Markov chain Monte Carlo approach includes the number of breaks K as a parameter in the model and explores both the model and parameter space jointly by ‘jumping’ between different numbers of breaks

- The proportion of time spent at each number of breaks is equal to the posterior model probabilities

- Using conjugate priors on the regression parameters β and σ allow us to marginalise them from the posterior and thereby explore the model space alone, greatly reducing the complexity of the algorithm.

Estimating parameter vector

- Sample parameters from FCs

$$\sigma^2_{ik} \mid \cdot \sim IG(\tilde{a}_k, \tilde{b}_{ik}),$$
$$\beta_{ik} \mid \cdot \sim MVN\left(\mu_{ik}, \Sigma_{ik}\frac{\sigma^2_{ik}}{\sigma^2_x}\right)$$

- σ^2_x is empirical variance of X from the full sample available at the time the forecast is made

- λ has been marginalised

Empirical Results

Empirical Results

- Forecast 30 portfolio returns and constructing the market forecast as weighted average with univariate regressions
  - Dividend-price ratio (aggregate)
  - T-bill
  - Term spread
  - Default spread
- Benefits of using panel to predict industry portfolio returns
  - Real-time break detection to exploit out-of-sample predictability

Data

- Monthly data from July 1926 - December 2015
- 30 industry portfolio returns (with and without dividends)
- 5×5 portfolio sorts on size and value and size and momentum
- Weights are constructed using two variables from French's website
  - Average firm size in portfolio
  - Number of firms in portfolio
- Returns in excess of risk-free rate and predictors from Goyal and Welch

Testing for cross-sectional dependencies

- Residuals e_{it} computed from OLS regressions for series $i = 1, \ldots, N$

$$e_{it} = r_{it} - \hat{\mu}_i - \hat{\beta}_i' X_t$$

- Estimate pairwise correlations to compute CD statistic of Pesaran (2004)

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

- Our approach reduces CD from 168.84 to 2.39 and average pairwise correlation from 0.74 to 0.13

- Any remaining correlations are likely to be weak and will not compromise inference in large panels ($N > 10$)

Models

- Unit-specific panel with pooled breaks (τ is common across units)

$$r_{it} = \mu_{ik} + \beta_{ik} X_{t-1} + \epsilon_{it}$$

- Benchmark I: Unit-specific panel with no breaks

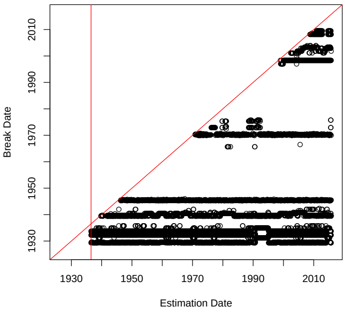
$$r_{it} = \mu_{i} + \beta_{i} X_{t-1} + \epsilon_{it}$$

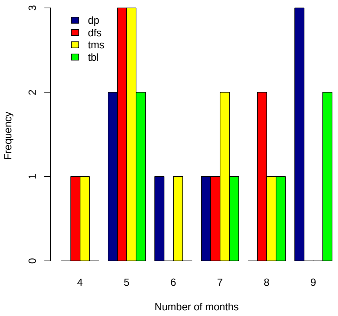
- Benchmark II: Time-series break model (Chib,1998) (τ_i is unit specific)

$$r_{it} = \mu_{ik} + \beta_{ik} X_{t-1} + \epsilon_{it}$$

- Benchmark III: Prevailing mean model

Real-time detection of breaks





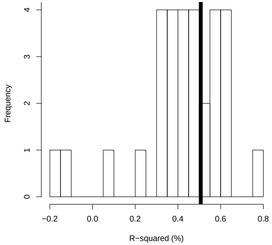
Forecast evaluation

- Evaluate the forecasting ability of each of the models relative to the benchmark model through the out-of-sample R^2 measure

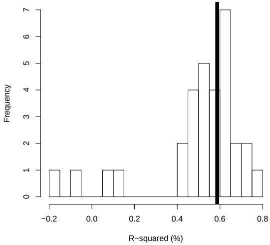
$$R^2_{OoS} = 1 - MSE_{Pbrk}/MSE_{bmk}$$

- MSE_{Pbrk}: MSFE for the heterogeneous panel break model
- MSE_{bmk}: MSFE for the benchmark
  - Positive R^2_{OoS} values suggest that the panel break model outperforms the benchmark
  - Statistical significance (Diebold and Mariano, 1995; Clark and West, 2007)

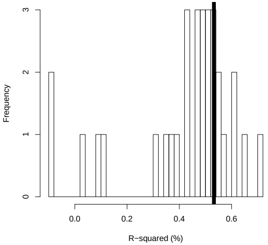
Histogram of R^2_{OoS} values across 31 cases



Hetero. no break panel



Time series break



Prevailing mean model

Out-of-sample forecasts: Diebold-Mariano tests

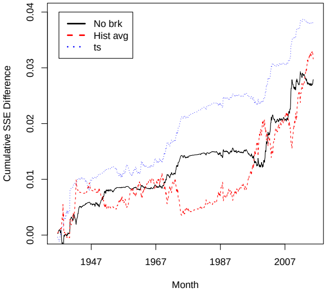
Dividend-price ratio

Predictor	DM				CW			
	$t < -1.64$	$-1.64 < t < 0$	$0 < t < 1.64$	$t > 1.64$	$t < -1.64$	$-1.64 < t < 0$	$0 < t < 1.64$	$t > 1.64$
No break panel								
dp	0	2	4	$25^{\dagger}$	0	2	2	$27^{\dagger}$
Industry prevailing mean								
dp	0	2	5	$24^{\dagger}$	0	2	3	$26^{\dagger}$
Time series break								
dp	0	2	3	$26^{\dagger}$	0	2	2	$27^{\dagger}$

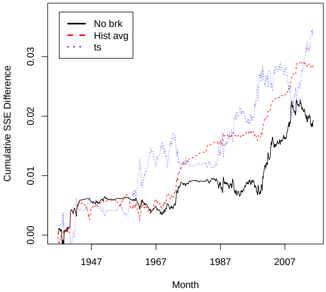
Panel model consistently beats benchmarks

- The cumulative sum of squared error difference is defined as

$$CSSED_{it} = \sum_{\tau=1}^{t} (\epsilon^2_{Bmk,i\tau} - \epsilon^2_{PBrk,i\tau})$$



(a) Market



(b) Telecommunications

Utility gains by portfolio

- At time t the mean-variance investor allocates a portion of his wealth to the portfolio at period $t+1$

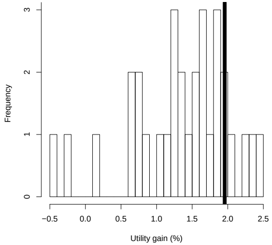
$$w_{i,t+1} = \frac{1}{\gamma} \frac{\hat{r}_{it+1|t}}{\hat{\sigma}^2_{t+1|t}}$$

- $\hat{\sigma}^2_{t+1}$ is five year rolling window of monthly stock returns to estimate the variance of stock returns (Campbell and Thompson, 2008)
- Moderate risk aversion ($\gamma = 3$)
- Investor realises an average utility of

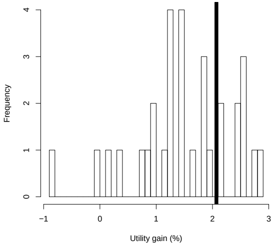
$$\hat{\nu}_{i,0} = \hat{\mu}_{i,0} - \frac{\gamma \hat{\sigma}^2_{i,0}}{2}$$

- Utility gain equals the difference multiplied by 1200

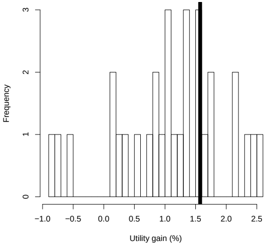
Histogram of ΔU values across 31 cases



Hetero. no break panel



Time series break



Prevailing mean model

Utility gains from allocating across industry portfolios

- Allocate wealth between r_f and risky portfolio constructed from 30 portfolios (Avramov and Wermers, 2006; Banegas et al, 2013)
- $R_{p,t+1}$: excess return on the risky portfolio at time $t+1$
- ω_t: portfolio weights
- Numerical methods used to compute ω_t that maximises

$$E[U(R_{p,t+1} \mid \gamma)] = r_{ft} + \omega_t' \hat{r}_{0,t+1} - \frac{\gamma}{2}\omega_t' \hat{\Sigma}_{0,t}\omega_t$$

- subject to $\sum_{i=1}^{N} \omega_{it} = 1$ and $\omega_{it} \in [0,1]$ for $i = 1, \ldots, N$
- $\hat{\Sigma}_{0,t}$ estimated using residuals from return model up to time t
- ω_t is plugged into utility function to obtain realized utility for time t
- Difference between utility values is annualized

Utility gain from allocating across portfolios

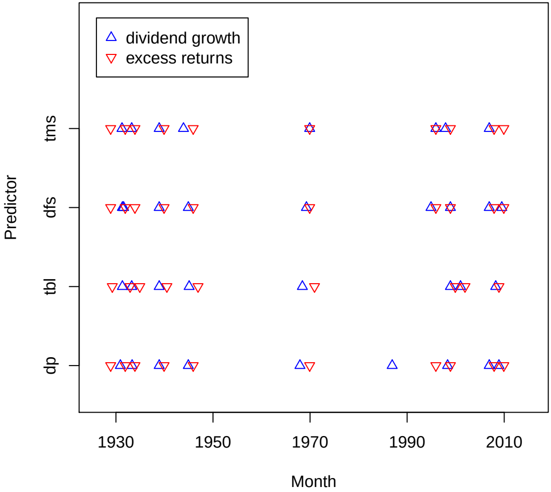
Predictor	Full sample			After breaks		
	hist avg	no brk	ts	hist avg	no brk	ts
dp	2.19	2.02	1.97	3.02	2.43	2.72
tbl	2.04	2.10	2.34	2.61	2.80	3.06
tms	1.99	2.42	1.86	2.21	2.57	2.37
dfs	2.02	1.92	2.29	2.89	2.53	2.72

Sources of breaks in return predictability

Sources of breaks in return predictability

- Return predictability could arise from two sources
  - time-varying risk premia
  - time-varying expectations of cash flows
- Cochrane (2008): little-to-no predictability in dividend growth
- Chen (2009), Binsbergen et al. (2010), Kelly and Pruitt (2013): some evidence of dividend growth predictability
- Disagreements about dividend growth predictability could be due to breaks in the dividend growth process
- Are breaks in return predictability linked to breaks in the dividend growth process?

Breaks in dividend growth



Is there a break risk factor?

Is there a break risk factor?

- Breaks reveal macroeconomic shocks
  - e.g. oil price shocks, financial crisis etc
- Are stocks more exposed to shocks priced in the cross-section?
- Produce forecasts for 7,299 stocks using panel model with and without breaks
- Break risk exposure proxied by mean squared forecast difference
- Sort stocks into quintile portfolios according to this measure

Is there a break risk factor?

Fama-Macbeth Regressions

Independent variable	Slope coefficients and (t-stats)
Break risk	0.59
	(4.60)
log(B/M)	0.30
	(5.19)
log(ME)	-0.08
	(-3.02)
PR1YR	0.57
	(3.17)

Is there a break risk factor?

Portfolio	r	α	MKT	SMB	HML
Low	0.26	-0.18	1.02	0.01	0.05
	(1.98)	(-2.04)	(22.21)	(1.43)	(3.93)
2	0.32	-0.06	0.98	0.03	0.13
	(2.19)	(-1.99)	(31.98)	(1.69)	(2.80)
3	0.44	-0.01	0.94	0.00	-0.09
	(2.25)	(-1.60)	(32.04)	(2.06)	(-1.21)
4	0.46	0.02	1.01	0.06	0.14
	(1.98)	(1.01)	(24.09)	(1.31)	(2.02)
High	0.53	0.17	0.96	-0.03	-0.03
	(2.58)	(2.04)	(23.26)	(1.99)	(-3.78)
High-low	0.27	0.35	-0.06	-0.04	-0.08
	(2.18)	(2.97)	(-1.05)	(-2.03)	(-1.55)

Break exposure and company characteristics

Portfolio	Size of break rank	s.d. across regimes Intercept	Slope	Volatility
		Size and Value		
SMALL.HiBM	1	0.067	0.141	0.088
ME1BE4	2	0.059	0.132	0.078
ME1BE3	3	0.058	0.129	0.078
ME1BE2	4	0.051	0.111	0.069
SMALL.LoBM	5	0.041	0.107	0.067
BIG.HiBM	20	0.032	0.052	0.054
ME5BE4	21	0.028	0.050	0.048
ME5BE3	23	0.024	0.044	0.039
ME5BE2	24	0.021	0.038	0.029
BIG.LoBM	25	0.019	0.030	0.024
		Size and Momentum		
SMALL.LoPRIOR	1	0.049	0.091	0.087
ME1PRIOR2	2	0.043	0.080	0.082
ME1PRIOR3	3	0.032	0.062	0.075
ME1PRIOR4	4	0.031	0.060	0.071
SMALL.HiPRIOR	5	0.027	0.056	0.052
BIG.LoPRIOR	20	0.029	0.042	0.046
ME5PRIOR2	22	0.020	0.036	0.042
ME5PRIOR3	23	0.012	0.025	0.038
ME5PRIOR4	24	0.009	0.022	0.026
BIG.HiPRIOR	25	0.008	0.020	0.022

Robustness checks

Results using other predictors

- Forecast with three other predictive variables
  - T-bill
  - Term spread
  - Default spread
- Improved forecasts ✓
- Statistically significant ✓
- Economically meaningful ✓

Conclusions

Conclusions

1. We propose a new approach to predict returns in the presence of model instability
   - Exploit information in cross-section to detect breaks in real time
   - Use economically-motivated prior to estimate magnitude of break
2. Demonstrate the usefulness of the method in an application to industry portfolio returns
   - Outperforms a range of competing benchmark specifications, including univariate time-series break model and a no-break panel model
   - Fast real-time break detection
   - Improvements to predictive accuracy are significant and economically meaningful
   - Results are robust to using a range of predictors
3. Break risk factor is priced in the cross-section