Technical Report
Locating changes in return-distributions for exchange traded funds

Liam Hayes
Colorado State University

Piotr Kokoszka
Colorado State University

Tim Kutta
Colorado State University

July 7, 2023

Summary

In this report, we investigate changes of return distributions for nine exchange traded funds (ETFs), each representing a different sector of the US economy (energy, industry, technology, etc.). Specifically, we focus on such changes in the wake of the Great Recession. We aim at identifying which sectors changed and when those changes occurred. This analysis provides a new understanding of where the crisis started and how it cascaded through the economy. Our main tool of statistical inference is the sectoral estimator, proposed recently by Kutta *et al.* (2023) in the context of virus data.

Keywords: change points, exchange-traded funds, mutlivariate functional data, financial crisis

Contents

1 Introduction

Returns of ETFs: An *Exchange Traded Fund* (ETF) is a basket of securities that is traded throughout the day on an exchange, similar to stocks. ETF prices and therefore returns are determined by supply and demand for shares in the fund as well as the underlying value of the holdings. Returns on these ETFs indicate the change in value and investor sentiment of the thing that the ETF is tracking. Put more simply, the return on an ETF represents the average change in value of the underlying securities (Ferri, 2009). If an ETF or basket of ETFs are set up to track important parts of the economy, returns on these ETFs can be a good reflection of the status of the economy.

The Great Recession: The *Great Recession* of 2008 was caused by the bursting of the housing bubble due to excessive lending of subprime mortgages to people who were not qualified for the given interest rates. As these mortgages began to default and become worthless, foreclosures increased and banks began struggling because a large portion of their assets were now worthless. As a result, liquidity of global financial markets decreased, which had a resounding effect on domestic and international economies. The Great Recession was the worst economic downturn since the Great Depression. During the Great Recession, GDP fell by 4.3% and unemployment rose from 5% to its peak of 10% in October 2009. In this report, we study the development of the Great Recession, by searching for structural breaks in return distributions of ETFs.

Inferring changes: Since the original work of Page (1954) *change point analysis* has become an important branch of statistical inference to investigate the stability of time series. Gromenko *et al.* (2017) recently proposed a test to detect changes in multivariate functional data (the focus of this report), with simultaneous changes for all component. Kutta *et al.* (2023) employed a similar strategy for asynchronous changes in a panel of random densities, to infer those components where the changes took place.

In this report, we modify this method, studying a panel of random densities for ETF returns. There are a three main differences between the approach taken in this report, compared to the original method on Kutta *et al.* (2023). First, we do not transform the densities into the Bayes space, but study them as L^2-function. Accordingly, the mean density is to be understood in a pointwise or L^2-sense, not as a Fréchet-mean in the Bayes space. Not moving to the Bayes space also removes the need to lower bound our densities and thereby eliminates a meta-parameter. Secondly, we do not impose a separable model, which is not required as Kutta *et al.* (2023) point out. Third, the focus of our economic analysis is not the early COVID pandemic, but the US economy around the Great Recession.

The structure of this report: This report consists of two parts. In Section 2 we introduce the statistical model and discuss the CUSUM statistic as a means to test for the presence of an abrupt change in the time series of distributions. In Section 3 we employ this method to infer in which components of the time series a change occurred. We will apply the resulting method to ETF values at the start of the Great Recession.

2 A statistical test for changes in return distributions

sec_2

2.1 Data and problem description

Data: We are looking at the SPDR sector ETFs during the Great Recession. These are a collection of ETFs that are set up to track 9 sectors of the economy. They track materials, energy, financials, industrials, technology, consumer staples, utilities, health care, and consumer discretionary. The data used in this report was investigated in a different context, e.g., by Kokoszka *et al.* (2023).

In the following, we refer by $s = 1, ..., S = 9$ to the different securities under considerations and denote by $t = 1, ..., T$ the trading day (where T is the total number of days under consideration).

Return distributions: We can split each trading day into consecutive intervals of time with equal length. In our case we chose $N = 39$ intervals (10 minutes each). For each sector s, day t and interval $i \in \{1, ..., N\}$ we compute the return

$$X(s,t,i) = \frac{P_{s,t,i} - P_{s,t,i-1}}{P_{s,t,i-1}} \text{ where } P_{s,t,i} \text{ is the price of } s \text{ at day } t \text{ on the end of interval } i.$$

We assume that the variables $X(s,t,i)$ are i.i.d. distributed across i, i.e. that short term price movements are unrelated. However, we do allow the return distributions to fluctuate from one day to another, where the underlying return distributions are *random densities*. A change in our economic time series has occurred if the mean return distribution (mean of the random densities) has changed at some point. In Figure 1 we display an estimated return distribution.

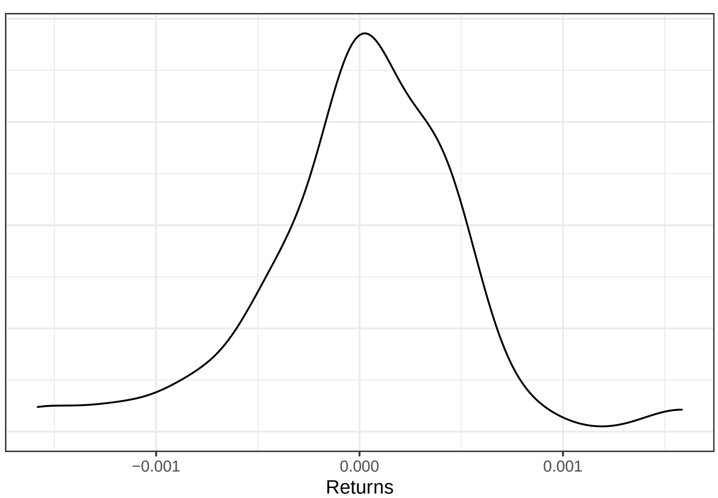


fig_1 Figure 1: Example of a distribution of returns for the Technology sector on 12/30/2010.

Our subsequent statistical inference will be based on kernel density estimators $f_{s,t}$ for all s, t that approximate the true return distribution $\mathbf{f}_{s,t}$ of $X(s,t,i)$. For computational

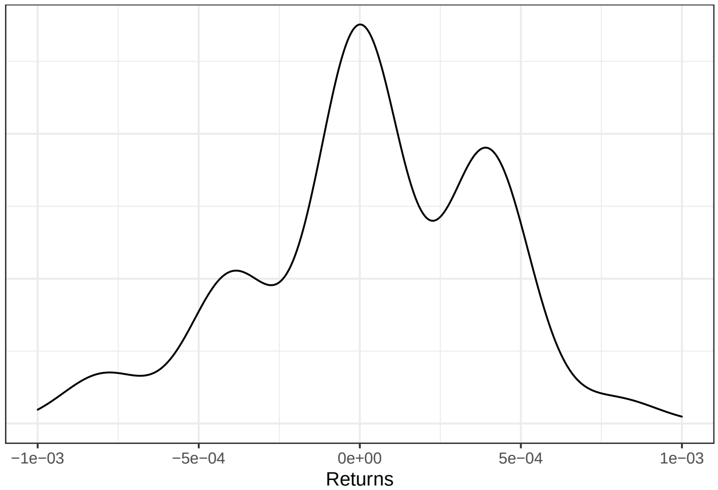


Figure 2: Distribution of returns with support truncated to the interval $[-0.001, 0.001]$

purposes, we discretize our density estimators by evaluating them on a fixed, equidistant grid with 50 gridpoints. The grid is laid over a support interval I, where for most densities, most of their probability mass is located. We found a reasonable choice for I to be from -0.001 to 0.001. [1]. See Figure 2 for a return density on our chosen support.

Inference for a change in the mean density: Let us formalize our previous considerations: For each sector s and trading day t we obtain an estimator $f_{s,t}$ that approximates the true, underlying return distribution $\mathbf{f}_{s,t}$. The estimator $f_{s,t}$ is then vectorized by evaluating it on the grid of $N = 50$ points on I. We refer to the vectorized version by $\overrightarrow{f_{s,t}} \in \mathbb{R}^N$ and to the the stacked vector (all sectors on one day) by $\overrightarrow{f_t} \in \mathbb{R}^{SN}$. We are interested in a change in the mean density. Here, the mean density at (s,t) evaluated in x can be understood as the pointwise mean of the density $\mathbb{E}\mathbf{f}_{s,t}(x)$. Since the kernel density estimators provide close approximations of the true densities $f_{s,t} \approx \mathbf{f}_{s,t}$, we can base our inference on the estimators in the following.

We assume that there exists at most one change in the mean density, i.e. there exist densities $\mu_s^{(1)}, \mu_s^{(2)}$ and a change point k_s^* s.t.

$$\mu^{(1)} = \mathbb{E}\mathbf{f}_{s,1} = \mathbb{E}\mathbf{f}_{s,2} = ... = \mathbb{E}\mathbf{f}_{s,k_s^*}, \quad \mu^{(2)} = \mathbb{E}\mathbf{f}_{s,k_s^*+1} = \mathbb{E}\mathbf{f}_{s,k_s^*+2} = ... = \mathbb{E}\mathbf{f}_{s,T}$$

and a change has occurred at k_s^* if $\mu_s^{(1)} \neq \mu_s^{(2)}$. We will ultimately be interested in deriving the set of all sectors where a change has occurred, or more formally to calculate a set estimator $\widehat{\mathcal{A}}$ for

$$\mathcal{A}^* := \{s : \mu_s^{(1)} \neq \mu_s^{(2)}\}.$$

Our estimator will follow the construction in Kutta *et al.* (2023), which guarantees that for large T it holds with probability ≈ 1 that $\mathcal{A}^* \subset \widehat{\mathcal{A}}$, while the risk of including sectors

[1]The code we used to explore these densities can be found in the file draw_hist_KDE.R: https://github.com/LiamCHayes/change-point-finance

where no change occurred is approximately equal to some user determined nominal level α (in our case always 5%). As a first step, we present a statistical test for the hypothesis of no change in a set of sectors $\mathcal{A} \subset \{1, ..., S\}$.

2.2 A test for no change in a set of sectors

Consider a set of sectors $\mathcal{A}$ (for example $\mathcal{A} = \{1, 2, 3\}$). We want to present a statistical test for a change in the mean density in any of these sectors. Our approach consists of two steps: First, we calculate a CUSUM statistic, second, we approximate its distribution under the null-hypothesis.

Test statistic:

i) **Vector concatenation:** Suppose $\mathcal{A} = \{s_1, ..., s_A\}$ (where always $s_1 < ... < s_A$ and $A = |\mathcal{A}|$ denotes the number of considered vectors). For each $t = 1, ..., T$ create the concatenated vectors

$$\overrightarrow{f_t^{\mathcal{A}}} := (\overrightarrow{f_{s_1,t}}, ..., \overrightarrow{f_{s_A,t}}).$$

Notice, that $\overrightarrow{f_t^{\mathcal{A}}}$ is of length $A \times N$.

ii) Create a function PS for **the partial sum process**, with input parameter t ($= 1, ..., T$) and region set $\mathcal{A}$ (subset of $\{1, ..., S\}$) defined as

$$PS(t, \mathcal{A}) := \sum_{r=1}^{t} \overrightarrow{f_t^{\mathcal{A}}}.$$

iii) Create the CUSUM-type statistic

$$CUSUM(\mathcal{A}) := \frac{1}{NAT^2} \sum_{t=1}^{T} \|PS(t, \mathcal{A}) - (t/T) \cdot PS(T, \mathcal{A})\|^2$$

where $\|\cdot\|$ denotes the common vector norm ($\|(1, 2)\|^2 = 1^2 + 2^2$).

p-values:

i) For a set of indices $\mathcal{A}$, **calculate the empirical covariance matrix**

$$\hat{C}^{(\mathcal{A})} := \frac{1}{2NAT} \sum_{t=2}^{T} [\overrightarrow{f_t^{\mathcal{A}}} - \overrightarrow{f_{t-1}^{\mathcal{A}}}] \cdot [\overrightarrow{f_t^{\mathcal{A}}} - \overrightarrow{f_{t-1}^{\mathcal{A}}}]^T.$$

ii) Calculate the **first** B **eigenvalues of** $\hat{C}^{(\mathcal{A})}$, say $\lambda_1^{(\mathcal{A})} \geq \lambda_2^{(\mathcal{A})} \geq ... \geq \lambda_B^{(\mathcal{A})}$. Set in our case as default $B = 30$.

iii) **Simulate the** p**-values** of the limiting distribution

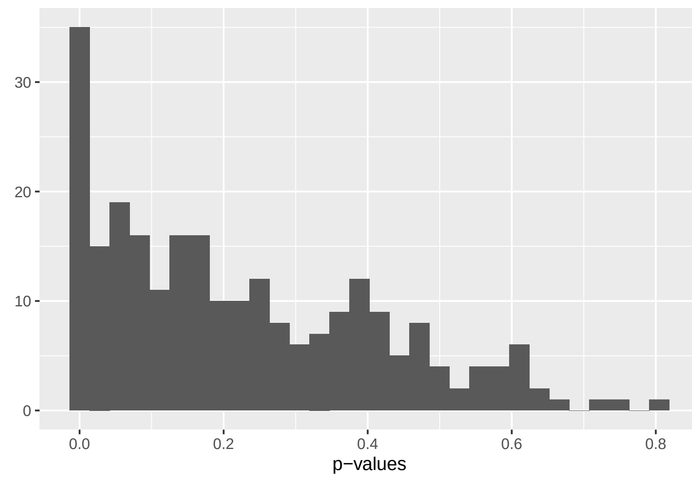


fig:10d_hist

(a) 10-Day Histogram

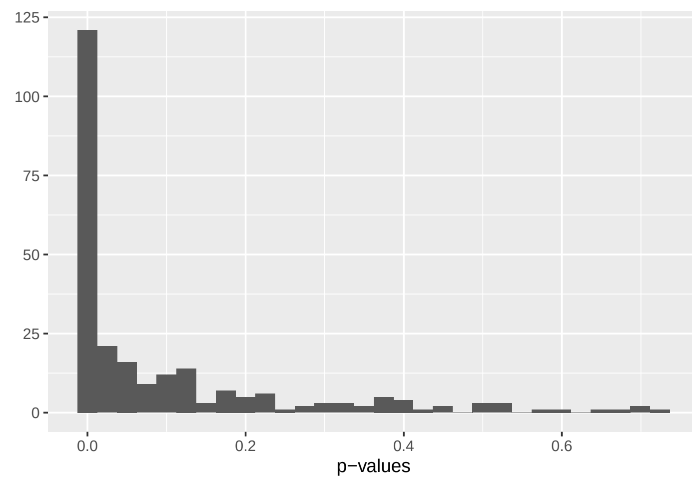


fig:30d_hist

(b) 30-Day Histogram

_comparison

Figure 3: To investigate the performance of our statistical test decision we investigated the distribution of empirical p-values over periods of $T = 10$ and $T = 30$ days, where the starting date was picked at random for our data set. As we can see, even for a period of about one month, the bulk of p-values are smaller than 5% hinting at non-stationarity.

- For $\ell = 1, ..., 1000$ do:
  * Create a vector $(I_1^{(\ell)}, ..., I_B^{(\ell)})$, where $I_1^{(\ell)}, ..., I_B^{(\ell)}$ are square integrated Brownian Bridges.
  * Calculate
$$v_\ell = \sum_{b=1}^{B} \lambda_b^{(\mathcal{A})} \cdot I_b^{(\ell)}$$
- Order the results in a vector s.t. $v_1 \leq ... \leq v_{1000}$.
- We now want to calculate the p-value of a number $r \geq 0$. For this purpose, take the v_i closest to r (minimizing $|r - v_i|$). Then, for this i
$$p^{(\mathcal{A})}(r) := 1 - \frac{i}{1000}.$$
- Calculate the p-value $p^{(\mathcal{A})} := p^{(\mathcal{A})}(CUSUM(\mathcal{A}))$. If $p^{(\mathcal{A})} < 0.05$ we reject the hypothesis of "no change in this set of sectors during the time frame $t = 1, ..., T$".

Examples: To better understand our test procedure, we ran the test with $\mathcal{A} = \{1, ..., 9\}$ and randomized starting points for different time periods. This repeated random sampling yields a distribution of p-values that shows how the test typically behaves for this type of data. Figure 3 shows the distribution of p-values of a 10 day and a 30 day time period with random start dates.

3 Locating changes

sec_3

We focus on locating changes during the Great Recession in the year 2008. This is a particularly interesting period because there were many major financial events such as monetary policy changes, stock market crashes, and changes in investor behavior that had resonating effects throughout the global financial system. There were 253 trading days during 2008 starting on January 2 and ending on December 30.

3.1 Identifying regions where a change has occured

Spatial identification

- For each $s = 1, ..., 9$ calculate the p-value $p^{(s)}$ for this region.
- Order the regions: We now reorder the regions into $s_1, s_2, ..., s_9$. The first region here is that with the smallest p-value, the second one with the second smallest p-value etc.
- Identification procedure:
  - Set $\widehat{\mathcal{A}} = \emptyset$
  - Calculate $p[s_1, ..., s_9]$. If $p[s_1, ..., s_9] \geq 0.05$ stop. If $p[s_1, ..., s_9] < 0.05$ update $\widehat{\mathcal{A}} = \{s_1\}$
  - Calculate $p[s_2, ..., s_9]$. If $p[s_2, ..., s_9] \geq 0.05$ stop. If $p[s_2, ..., s_9] < 0.05$ update $\widehat{\mathcal{A}} = \{s_1, s_2\}$
  - Calculate $p[s_3, ..., s_9]$. If $p[s_3, ..., s_9] \geq 0.05$ stop. If $p[s_3, ..., s_9] < 0.05$ update $\widehat{\mathcal{A}} = \{s_1, s_2, s_3\}$
  - ...
  - Calculate $p[s_9]$. If $p[s_9] \geq 0.05$ stop. If $p[s_9] < 0.05$ update $\widehat{\mathcal{A}} = \{s_1, s_2, s_3, ..., s_9\}$. Stop
- Return $\widehat{\mathcal{A}}$

For any $s \in \widehat{\mathcal{A}}$ calculate the estimated time of a change

$$\hat{t}_s := \text{argmax}_t \| PS(t, \{s\}) - (t/T) \cdot PS(T, \{s\}) \|^2.$$

Inference for the Great Recession: Since there is such a high frequency of changes, we decided to look at the Great Recession month by month. For each month in 2008, the spacial identification process was repeated. Here are the results of months where sectors showed changes:

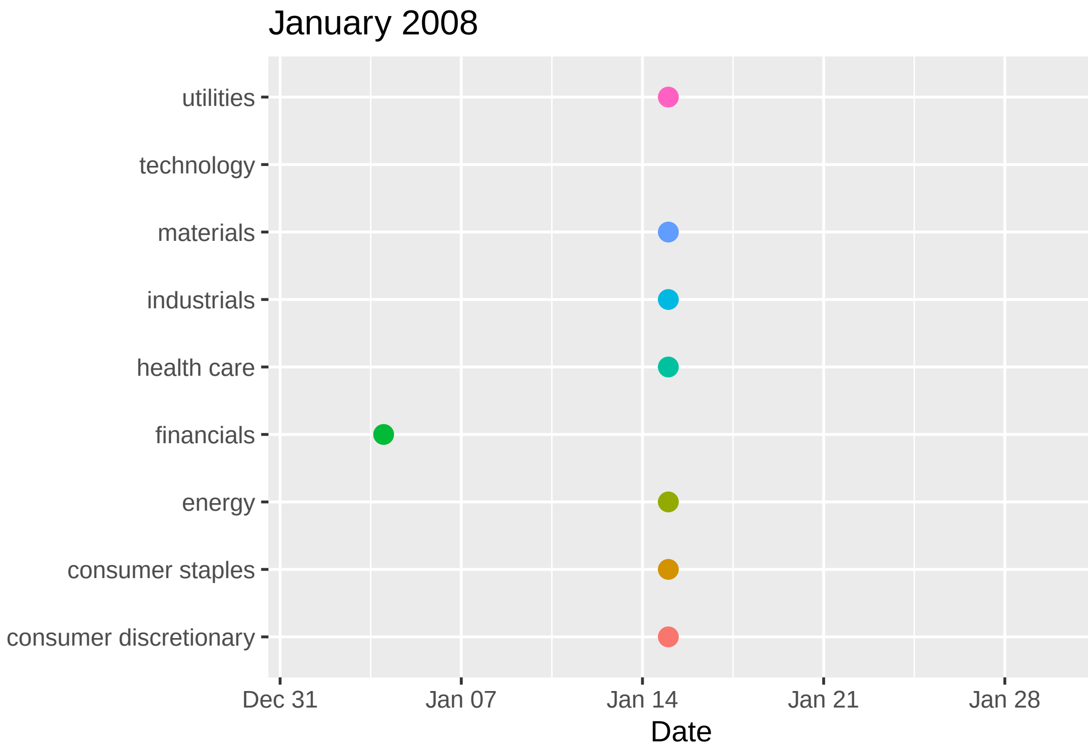


fig:Jan2008

Figure 4: Timeline of changes in January 2008

1. **January 2008:** In January, home sales fell to the lowest level in 10 years after the housing boom in 2006. On January 22, the FOMC lowered the Fed funds rate to 3.5%, and again to 3% a week later. Foreclosure rates spiked.

   The sectors that showed a change in distribution in January were materials, energy, industrials, consumer staples, utilities, health care, consumer discretionary, and financials.

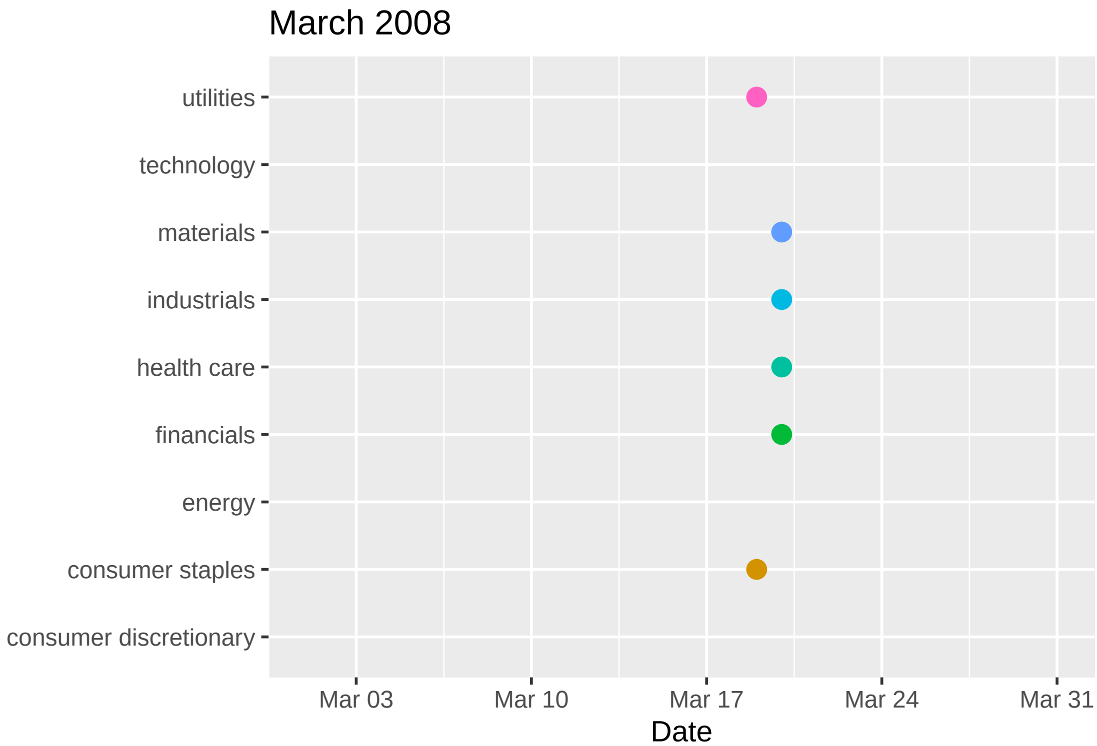


fig:Mar2008

Figure 5: Timeline of changes in March 2008

2. **March 2008:** In March, the Fed started bailouts and aggressive expansionary monetary policy. March 11 the Fed started bailing out bond dealers, March 17 they announced they would guaruntee Bear Sterns' bad bank loans, and March 18 the Fed Funds Rate was lowered by %0.75 to %2.25

   The sectors that showed a change in March were utilities, health care, financials, industrials, materials, and consumer staples.

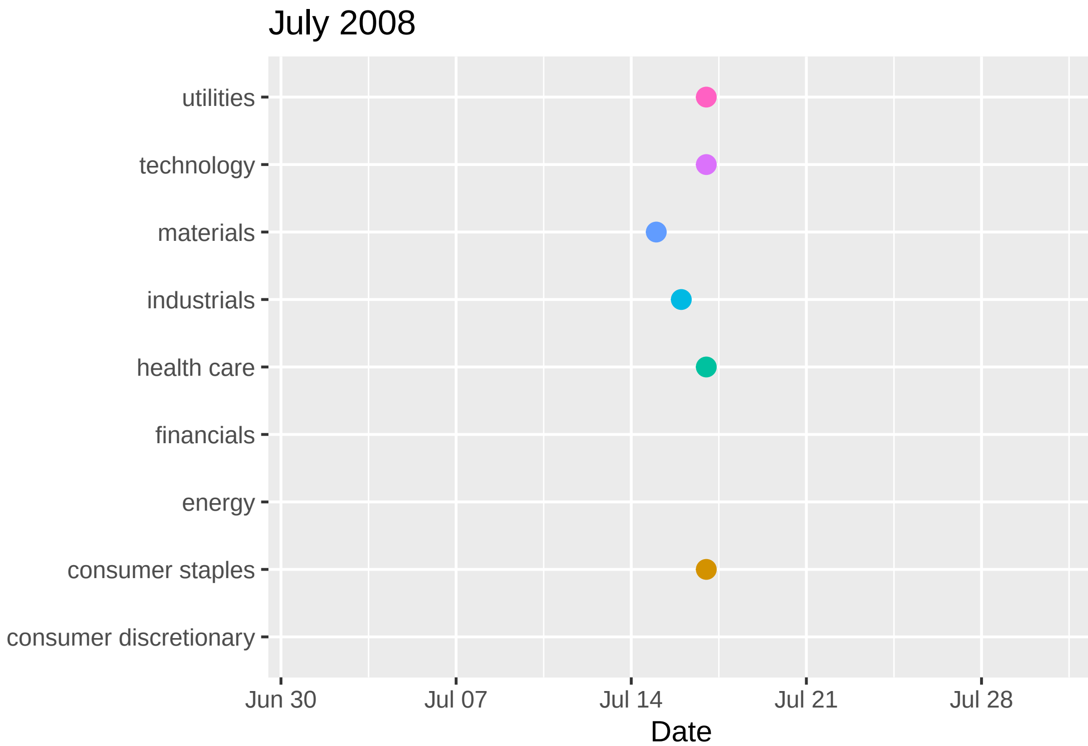


fig:Jul2008

Figure 6: Timeline of changes in July 2008

3. **July 2008:** On July 11, IndyMac bank failed. At this time, the FDIC only insured deposits up to $100,000. Multiple sectors showed a change directly after July 11. On July 30, Congress passed the Housing and Economic Recovery Act. This act created a new regulator for Fannie May and Freddie Mac and allowed for more loan guarantees, housing grants, and housing tax breaks.

   The sectors that showed a change in July were materials, industrials, utilities, consumer staples, health care, and technology.

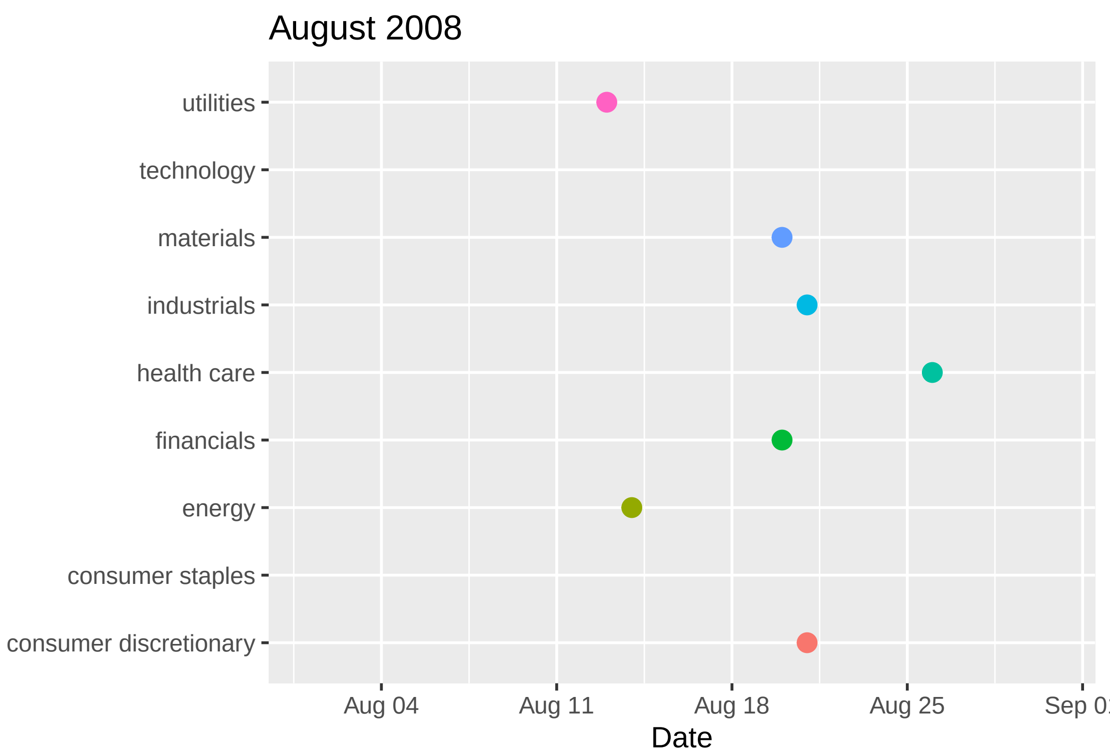
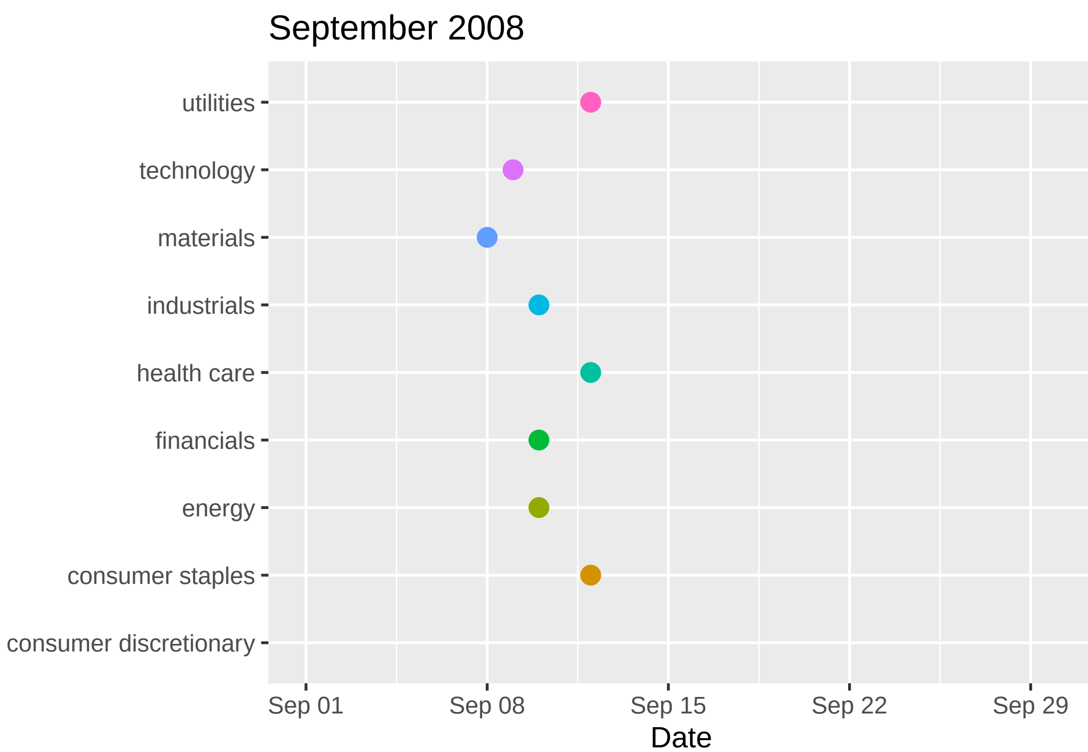


fig:Aug2008

Figure 7: Timeline of changes in August 2008

4. **August 2008:** No notable financial events happened in August, but sectors showed changes. This could be a result of the Housing and Economic Recovery Act being passed on July 30.

   The sectors that showed a change in August were materials, energy, financials, utilities, consumer discretionary, industrials, and health care.



g:Sept2008

Figure 8: Timeline of changes in September 2008

5. **September 2008:** The new regulator for Fannie Mae and Freddie Mac, the Federal Housing Finance Agency, placed Fannie and Freddie under conservatorship on September 7. On September 15, Lehman Brothers declared bankruptcy and triggered a global panic; The Dow fell 504 points and investors fled to safety by buying bonds. Oil prices also fell. Interestingly, sectors did not show changes after this event but showed changes after the conservatorship announcement on September 7. The rest of September continued the trend of bad news and bear markets culminating in a global stock market crash on September 29.

   The sectors that showed a change in September were materials, energy, financials, industrials, technology, consumer staples, utilities, and health care.

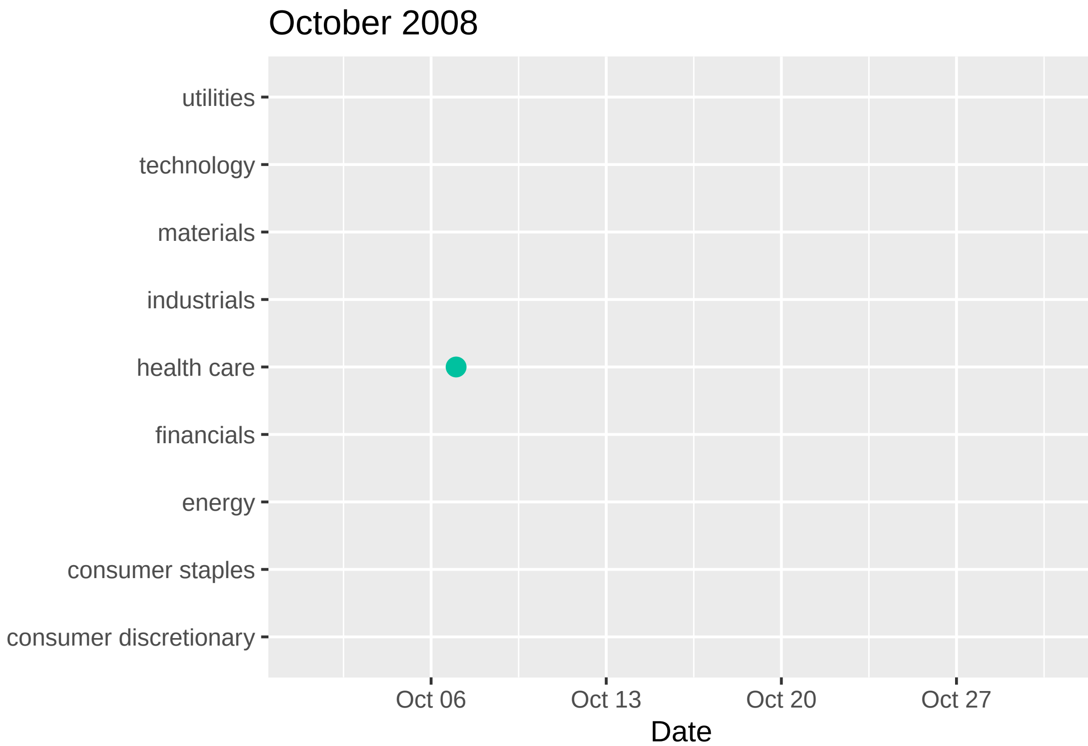


fig:Oct2008

Figure 9: Timeline of changes in October 2008

6. **October 2008:** Congress passed the $700 billion bank bailout bill on October 3 in an attempt to inject capital into the paralyzed financial system. Despite this attempt, global stock markets crashed again on October 6. The Fed and other central banks continued to try to unfreeze the economy throughout the month.

   The sector that showed a change in October was health care.

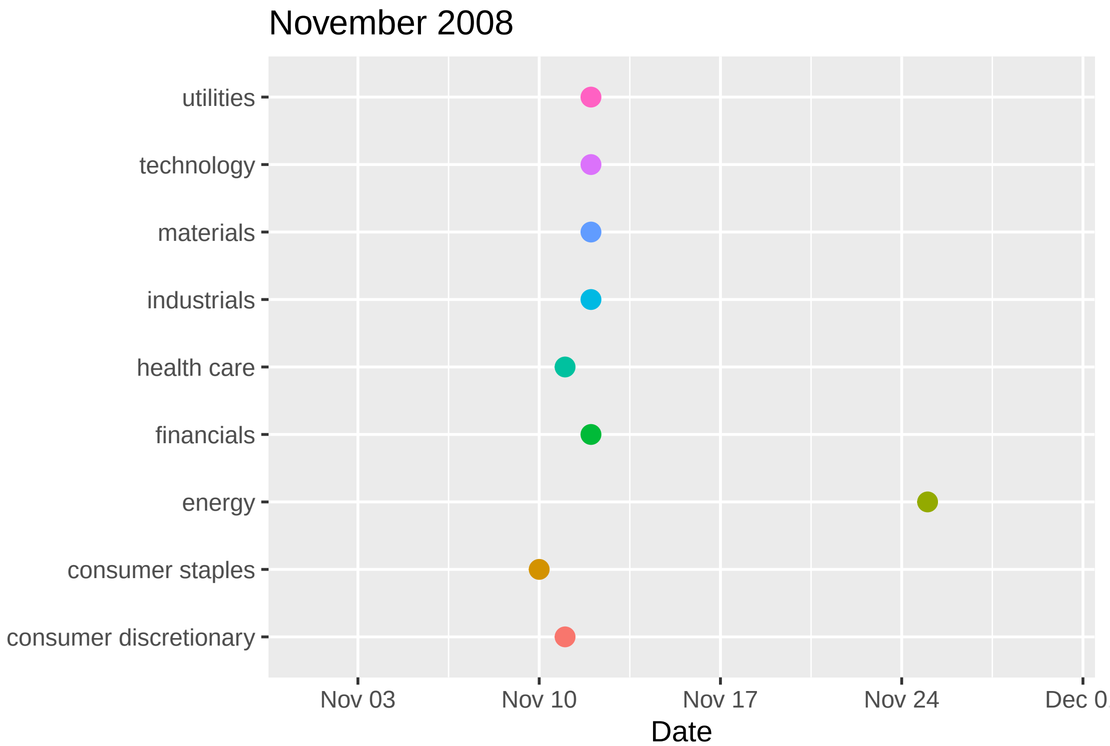


fig:Nov2008

Figure 10: Timeline of changes in November 2008

7. **November 2008:** More damage control from the government throughout the month.

All sectors showed a change this month.

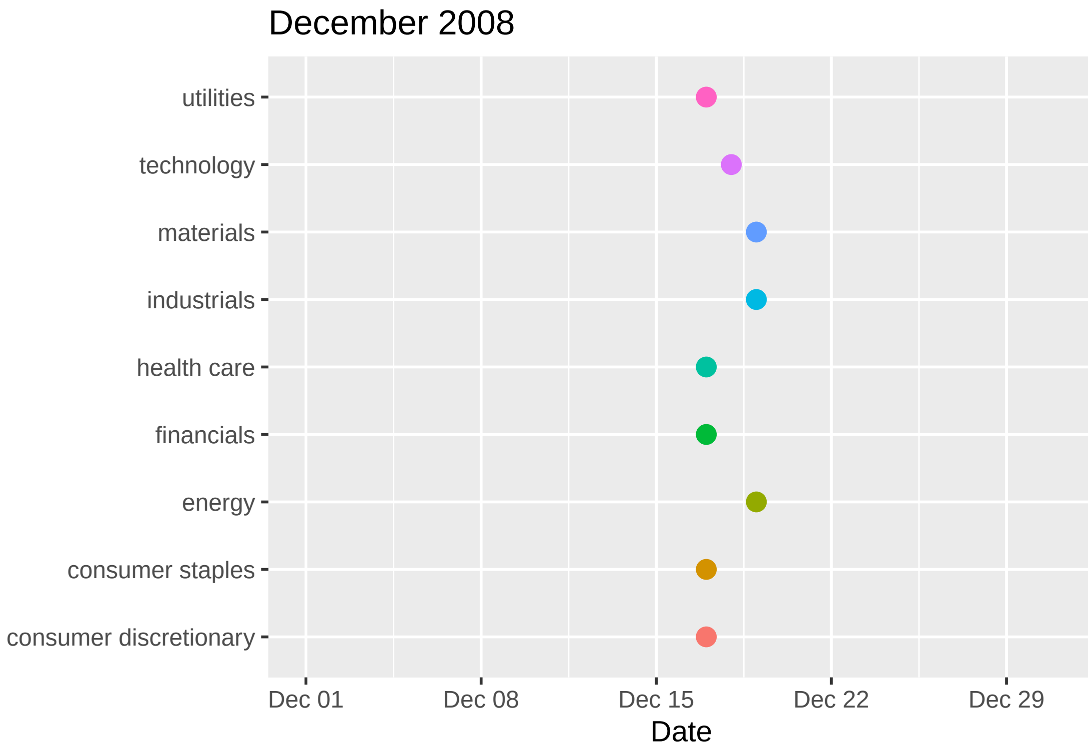


fig:Dec2008

Figure 11: Timeline of changes in December 2008

8. **December 2008:** Damage control continued in December and things started to get better. The recession officially lasted until June 2009, but the economy steadily recovered from here.

   All sectors showed a change this month.

References

Ferri:2009 Ferri, Richard (2009). *The etf book.* John Wiley and Sons, Inc.

mherr:2017 Gromenko, O., Kokoszka, P. and Reimherr, M. (2017). Detection of change in the spatiotemporal mean function. *Journal of the Royal Statistical Society (B)*, **79,** 29–50.

:wang:2023 Kokoszka, P., Mohammadi, N., Wang, H. and Wang, S. (2023). Functional diffusion driven stochastic volatility model. *arXiv:2305.04112*.

addad:wang Kutta, T., Jach, A., Kokoszka, P., Haddad, M. and Wang, H. (2023). Detection of changes in a panel of densities with application to viral analysis. *to appear*.

page:1954 Page, E. S. (1954). Continuous inspection schemes. *Biometrika*, **41,** 100–115.